# Which Firms Gain from Digital Advertising?

## Evidence from a Field Experiment *

Weijia Dai
Purdue University

Hyunjin Kim
INSEAD

Michael Luca
Harvard Business School
& NBER

January 2023

**Abstract**

Measuring the returns of advertising opportunities continues to be a challenge for many businesses. We design and run a field experiment in collaboration with Yelp across 18,294 firms in the restaurant industry to understand which types of businesses gain more from digital advertising. We randomly assign 7,209 restaurants to freely receive Yelp's standard ads package for three months. The scale of the experiment gives us a unique opportunity to assess the heterogeneity in advertising effectiveness across a variety of business attributes. We find that restaurants that receive advertising on Yelp observe on average a 7-19% increase in a wide range of purchase intention outcomes, as well as a 5% increase in customer reviews. We find that gains are heterogeneous across firms, with independent and higher-rated businesses observing larger gains, as well as those with more reviews and higher pre-experiment organic traffic.

*We thank Susan Athey, Rodrigo Belo, Garrett Johnson, Randall Lewis, Sridhar Moorthy, Jesse Shapiro, and Catherine Tucker and participants at the NBER Summer Institute, Conference on Digital Experimentation, Advances with Field Experiments Conference, and International Industrial Organization Conference for valuable comments. We thank Yelp for providing engineering support for this experiment. Dai, Kim, and Luca have consulted for tech companies, including Yelp, but their compensation and ability to publish did not depend on the results of this paper. Please contact Daisy Dai (dai196@purdue.edu), Hyunjin Kim (hyunjin.kim@insead.edu), or Mike Luca (mluca@hbs.edu) for correspondence. Authors are listed alphabetically.

# 1 Introduction

In recent years, digital advertising has been on the rise, exceeding expenses on television advertising and accounting for over $107 billion of spending in the United States in 2018 (Ha, 2019). This rise has made more data available for companies to assess their advertising decisions. Despite this availability of more granular data, estimating the returns to digital advertising remains difficult, leading many firms to overspend (Blake et al., 2015).[1] In principle, advertisers can conduct randomized experiments to measure the effect of advertising campaigns, but experiments are considered costly and sometimes infeasible for smaller businesses (Kohavi et al., 2020; Campbell et al., 2021).[2] While recent estimates using field experiments in collaboration with a few large advertisers have significantly advanced our understanding of the gains from advertising (Nelson et al., 2020), studies have found widely varying estimates,[3] highlighting the importance of understanding the heterogeneity and the types of businesses that are more or less likely to benefit.

This paper evaluates the impact of digital advertising using a field experiment on Yelp across more than 18,000 restaurants. Two features of our experiment are worth noting. First, the scale of our experiment spanning thousands of firms within an industry and the rich set of business attributes observed provides us with a unique opportunity to understand predictable heterogeneity in the effectiveness of ads. Beyond informing theory, our findings provide useful insights for practice, as many businesses lack the power to run experiments of their own. Second, most of the restaurants on Yelp did not advertise at the time of the experiment, and our sample consists of only non-advertisers. The sample thus provides us with an unusual opportunity to assess advertising effects for a broader set of businesses, building on prior literature that has generally analyzed advertising effectiveness for those who have already chosen to advertise (Shapiro et al., 2021).

We design and run our field experiment in collaboration with Yelp, the largest reviews and search engine platform for local businesses in the US. At the time of the experiment, businesses could purchase a standardized search advertising package at a fixed monthly rate, which guaranteed a minimum number of advertising impressions each month. We use this standard package as our

[1] Endogeneity in the advertising funnel and strategic behavior of the participants in the digital advertising ecosystem can severely bias estimates of advertising effectiveness (Lewis et al., 2011). Gordon et al. (2021) and Choi et al. (2020) offer excellent surveys of the challenges in digital advertising attribution. Johnson et al. (2022) highlight additional difficulties in tracking advertising gains in the new era of consumer privacy protection.

[2] Large sample sizes, often at the level of millions of impressions, are needed to estimate the impact of advertising with sufficient precision (Lewis and Rao, 2015; Athey and Imbens, 2016), limiting the use of experiments to large campaigns.

[3] Blake et al. (2015) ran experiments using geographic pausing of ads to assess the effectiveness of Google search advertising for eBay and found a negative return on advertising spend while Coviello et al. (2017) and Simonov et al. (2018) found different results for firms not as prominent as eBay.

treatment and randomize at the business level. Across a sample of 18,294 businesses that did not previously advertise on Yelp, we randomly assign 7,209 restaurants to freely receive these standard ad packages for three months, sending out roughly 500 million total paid search impressions. We observe business-level outcomes over a 12-month period before and after treatment, allowing us to assess the causal effects of advertising on page views of the business's Yelp page, a standard measure for advertising effectiveness (Graepel et al., 2010; McMahan et al., 2013), as well as additional measures of consumer intentions to visit the restaurant. We explore heterogeneity in advertising effects by leveraging the platform's rich data on business attributes.

We find that on average, advertising lifts the number of page views, purchase intentions, and reviews. Advertising results in a 19% increase in restaurants' page views; a 7%–14% increase in purchase intention outcomes further down the funnel, such as calling the business, viewing map directions, and browsing the restaurant's website; and a 5% increase in the number of customer reviews. We also find that these advertising effects dissipate immediately on the outcome measures we track once the advertising packages are suspended, suggesting that advertising may only temporarily raise customer demand.

We find that advertising effects are heterogeneous across firms. We explore heterogeneity along different restaurant attributes measured at the beginning of the experiment—the restaurant's chain status, average review rating, organic traffic, number of reviews, age, and price level—and find important heterogeneity along three key dimensions. First, national chain restaurants gain less from advertising than independent and local chain restaurants with similar attributes. The latter also observe larger gains when they are newer or have higher ratings, which do not vary as much for national chains. Second, larger advertising gains accrue to restaurants with more reviews and higher baseline traffic on the Yelp platform. Third, advertising gains tend to be larger for more expensive restaurants, except for the most expensive ones. Taken together, our results are consistent with the role of search advertising, which more prominently places a business into a user's consideration set and provides information about the advertising business. In this context, advertisements essentially move a product to the top of visibility in a setting that otherwise has many products or firms being considered.

This paper relates to several streams of literature. It most directly contributes to the growing literature evaluating the effectiveness of digital advertising using experiments.[4] Many papers that

[4] Many papers have used randomized experiments to estimate the effectiveness of digital ads. For examples, Goldfarb and Tucker (2011), Lewis and Reiley (2014), Johnson et al. (2014), Sahni (2015), Johnson et al. (2017).

estimate the advertising effect using experiments often focus on one large company, and have found a wide range of estimates for the returns to advertising.[5] Some recent papers have also looked at effectiveness across different companies. Simonov et al. (2018) analyze the effectiveness of search advertising with brand keywords for thousands of firms and find large differences in returns depending on competition for the focal brand name keyword.[6] Rao and Simonov (2019) look at the impact of experimental results that gained wide publicity in the media (Blake et al. (2015)), but find no evidence of firms using experimental variations to assess returns to ads.[7] Our experiment contributes to this field by designing a large-scale field experiment that covers advertising campaigns for thousands of businesses with varying quality, age, and brand equity—allowing us to provide a detailed and managerially relevant cut of heterogeneous advertising effects across firms. Our experiment is also compelling in being able to turn advertising campaigns entirely on or off for these businesses, allowing us to see overall effects, and provide novel results on effects for a population of businesses that had not been previously advertising.

Our paper also relates to literature on different channels through which advertising influences demand (Nelson, 1974; Bagwell, 2007). Many papers have focused on two classic roles of advertising, informative and persuasive (Jeziorski and Moorthy, 2017; Sahni and Nair, 2020). Lovett and Staelin (2016) further distinguish two different roles of informing consumers: helping consumers learn their match value with the firm (the "informing" effect) and reminding consumers of the existence of the product by increasing the salience of the product in a consumer's memory (the "reminding" effect).[8] Our paper provides results consistent with these mechanisms, and closely relates to other work in similar settings exploring these mechanisms. Sahni and Nair (2020) provide evidence on the signaling effect of advertising using a field experiment on a mobile search and review platform similar to ours. Their experiment randomizes whether a paid listing disclosed its advertiser status, keeping its ranking position constant. They find that this disclosure increased calls to the restaurant, and that the effect was stronger for restaurants with fewer ratings and users from other cities—i.e., when uncertainty about the restaurant quality was larger. Biswas (2020) explores the role of providing

[5]For example, Blake et al. (2015) find that Google search ads have no measurable effects on demand for eBay while Coviello et al. (2017) find starkly different results from the search advertising experiment for Edmunds.com.

[6]Kalyanam et al. (2018) conduct a meta-analysis of 15 experiments on Google search ads involving 13 US multichannel retailers and find a positive return on ads spent but negative returns on investment, on average.

[7]Some recent progress has been made in the evaluating methods of ads effectiveness, such as designing less costly and more effective experiment methods (Johnson et al., 2017b; Sahni et al., 2019; Feit and Berman, 2019) and using results from experiments to benchmark non-experimental estimation methods based on observational data (Gordon et al., 2019, 2022).

[8]Lovett and Staelin (2016) found that paid media (advertising) mainly operates through reminding in the context of TV viewing choice.

specific information in ads (in this case, promotion amounts) over and above brand elements, and shows that adding explicit discount information can induce more consumers to search for the product but result in lower conversion conditional on searching. In comparison, our experiment varies whether the business advertises at all to understand the overall effects of advertising and its heterogeneity, disclosing that it is a paid ad and moving it to the top ranking position while keeping the information provided in the listing constant. Consistent with Sahni and Nair (2020), we find that newer firms gain more leads from advertising. However, we also find that gains are larger for firms with more reviews and higher organic traffic—and thus more likely to have higher quality, raising the possibility that effects other than signaling (e.g., improving consideration as Lovett and Staelin (2016) suggest), may be important as well in this setting.

Finally, our paper also provides new insights into the role of advertising when quality information exists, showing the importance of advertising in raising consumer awareness. While previous research such as Hollenbeck et al. (2019) has shown that ads outside the review platform are seen as substitutes for poor quality by firms, we find that they can serve as complements for high quality.[9] Essentially, ads on the platform provide a way to stand out even for already prominent restaurants, by rising to the top of consumer awareness among a list of potential choices.

# 2 Experimental Design

## 2.1 Context

Yelp is an online platform that allows consumers to search, browse, and review local businesses around the world. As of 2015, it had roughly 163 million monthly unique visitors. Yelp displays paid search ads when users search for local businesses, which constitute one of its main revenue streams. Its search function works similarly to other search engines. When a consumer enters a search for a business type or keyword (e.g., "italian") in the location she is looking for, the results page returns a list of businesses that best match the search query, with links to their pages on Yelp.

For each search, Yelp allocates a single slot for the paid search ad, which is listed at the top of the results page followed by organic search results (see Figure 1). Yelp determines which advertising business it shows in the paid search slot based on relevance and displays a prominent yellow flag labeled "Ad" next to the name of the business to distinguish it from organic results. If there is no

[9]Our paper also relates to the growing literature on firm strategy for reviews and review platforms (Mayzlin et al., 2014; Luca and Zervas, 2016; Zervas et al., 2017).

advertising business relevant to the consumer's search query, no paid search link appears on the result page. At the time of the experiment, Yelp's paid search ads were sold through standard monthly advertising packages, where businesses paid a $400 monthly fixed fee. Yelp managed the placement of ads in all three channels where searches were conducted—web browser, mobile website, and Yelp's mobile app. It also guaranteed a minimum number of paid search impressions for the advertiser.

### 2.2 Sample, Treatment, and Randomization

We run our experiment on US restaurants listed on Yelp. This represents the largest business category on Yelp, providing us with a large sample size to identify heterogeneity across business attributes. Furthermore, this category is representative of the entire population of restaurants as measured by official government statistics (Glaeser et al., 2022). Yelp also experiences a high volume of search in this category, which, combined with the low rate of advertising in 2015, provides us with a sufficient unused capacity of search sessions to run advertising experiments without impacting Yelp's existing advertisers.

As treatment, we provide Yelp's search advertising packages as is. We select a random set of eligible restaurants to be treated and provide them with the standard advertising package for free over a three-month period (from August 1 to October 31, 2015). All treated restaurants were advertised.

To define our sample, we first filter out businesses that are not actively operating and not qualified to advertise.[10] These requirements seek to screen out businesses with fraudulent information. Next, we exclude restaurants that have purchased advertising in the six months before the experiment to prevent having active advertisers opting in and out of advertising during the experiment period. Active advertisers accounted for a very small fraction of restaurants when we conducted the experiment, so excluding them has a negligible impact on representativeness.

We conduct stratified sampling in four subsamples of restaurants, which span across a wide range of business types and the outcomes we can observe. The first subsample (306 firms) consists of restaurants on Yelp Reservations, a reservation system hosted by Yelp that allows consumers to directly reserve on the restaurant's Yelp page. We observe the number of reservations made for these restaurants. The second subsample of restaurants are those on OpenTable (5,740 firms), another

[10] To be eligible for ads, a business must have verified basic information on their listing (e.g., an address and a phone number) and at least one photo contributed by a user or its owner. A minimum number of "trusted" three-star or better reviews—not flagged by Yelp's algorithm as fake—is also required.

reservation booking platform, which includes many mid-range to high-end full-service restaurants, similar to Yelp's small set of active advertisers. The third subsample consists of restaurants that have signed up to use EAT24, Yelp's delivery and take-out ordering service (5,867 firms), which enables us to observe the number orders placed through Yelp. The last subsample consists of restaurants not partnered through systems like Yelp Reservations, OpenTable, or EAT24, which are in the state of Washington (6,381 firms). We choose these restaurants because we can link them to historical revenue data that we have obtained from the Washington State Department of Revenue. The distribution of restaurant attributes across different subsamples is shown in Appendix A Table A3, and we discuss subsample treatment effects in Appendix G.

We obtain a total sample size of 18,294 restaurants, which we randomly assign to treatment at the restaurant level within each subsample to achieve better balance between treatment and control groups within each subsample (Imbens and Rubin, 2015). It also allows us to use different treated ratios across subsamples to avoid over-treating a local market, which can potentially lead to large general equilibrium effects.[11] We treat 50% of the restaurants in subsamples 1 to 3 and 20% of the restaurants in subsample 4 (see Table A1), as subsample 4 represents a larger fraction of restaurants in local markets.[12]

Our samples are balanced, with no significant differences in outcomes and characteristics between treated and control restaurants before the experiment (Table A2). We also plot restaurant attributes across treated and control firms in Appendix A Figure A1, which shows similarity in the distribution of attributes across both groups.

## 2.3 Outcome Measures

To assess the effectiveness of advertising, we track an array of purchase intention measures along the advertising conversion funnel. We observe these measures aggregated at the business-month level between January and December 2015. We first track the number of page views of a restaurant's Yelp page, which is a standard measure of advertising effectiveness used by the online advertising industry. This is the most direct measure of the effect of advertising: paid search promotes restaurants' visibility during customer searches, leading to an increase in traffic to the restaurant's Yelp page.

[11] If a high fraction of restaurants appears in search ads within a local market, search ads may become less effective, biasing down our estimate, or it could cause negative spillovers to control restaurants through business stealing, biasing our estimate upwards.

[12] Subsample 4 consists of over 6,000 restaurants in Washington state while the total number of restaurants on Yelp is around 200,000 in mid-2015. We treat 20% of subsample 4 so that the mean fraction of restaurants treated is 2.3% and the median 1.4% within a zip code.

Directly attributing paid page views to the effect of search advertising causes bias since it ignores the substitution of organic page views that the restaurant would have earned without advertising (Blake et al., 2015; Chesnes et al., 2017). Since we observe organic page views, we report the degree of the substitution effect between paid and organic page views.

We also track customer interactions with call-to-action (CTA) buttons on the business's Yelp page to get a better sense of the effects of advertising further down the purchase funnel. The clicks on CTA buttons, also referred to as "leads," allow customers to request map directions to the restaurant, directly call the restaurant, and visit the restaurant's own website outside Yelp. We also track the number of reviews written on the restaurant as an approximation for demand. For subsets of the experiment samples, we observe sales in online channels: the number of reservations for restaurants partnered with Yelp Reservations (subsample 1) and the number of take-out orders for restaurants partnered with Yelp EAT24 (subsample 3).

## 3 Empirical Specifications

We estimate the average treatment effect (ATE) using regressions similar to a difference-in-differences (DID) specification. Since the treated ratios differ across stratified subsamples by design, we modify the typical DID specification accordingly (Imbens and Rubin, 2015). The ATE estimates from this modified regression are equivalent to the weighted sum of subsample ATEs using subsample sizes as weights. Let the dummy variable $T_{\iota t}$, $\iota \in \{0, 1, 2\}$ denote whether month $t$ is in the period before, during, or after the experiment; the dummy variable $S_{si}$, $s \in \{1, 2, 3, 4\}$ denote the subsample that restaurant $i$ is in; and $\omega_i$, $\omega \in \{0, 1\}$ denote whether restaurant $i$ is assigned to the control or treatment group. We use the following specification to estimate the ATE:

$$\begin{aligned} y_{it} = &\beta_0 + \sum_{\iota=1}^{2}\sum_{s=2}^{4}\gamma_{s\iota}T_{\iota t}S_{si} + \sum_{\iota=1}^{2}\alpha_\iota\frac{1}{\lambda_1}\omega_i T_{\iota t}S_{1i} + \sum_{\iota=1}^{2}\sum_{s=2}^{4}\beta_{s\iota}(S_{si} - \frac{\lambda_s}{\lambda_1}S_{1i})\omega_i T_{\iota t} \\ &+ \mu_i + \delta_t + \epsilon_{it}, \end{aligned} \tag{1}$$

where $\lambda_s$ is the share of restaurants in subsample $s$. In equation (1), $\alpha_1$ and $\alpha_2$ consistently estimate the ATE during and after the experiment; $\beta_{s\iota}$ ($s = 2, 3, 4$; $\iota = 1, 2;$) respectively estimate the treatment effects for subsamples 2–4 during and after the experiment. The cross product terms $T_{\iota t}S_{si}$ control for subsample-specific time trends, and $\mu_i$ and $\delta_t$ control for restaurant and month fixed effects. We extend this equation to identify heterogeneous treatment effects, and explain the intuition underlying this specification in Appendix B.

Since the absolute scale of the effect contains proprietary information, we report ATE in standardized levels and in percentages. The outcomes are standardized by their respective full-sample standard deviations before the experiment. For ATE in percentages, we first calculate the percentage effect within each subsample. This is done by dividing subsample treatment effects by their control sample means during each period. We then take the weighted average of subsample percentage effects using sample sizes as weights. Since both the treatment effects and the denominators are random variables and may be correlated with each other, we calculate the confidence interval of the percentage effects using bootstrap.

# 4 Results

## 4.1 Average Treatment Effects

Table 1 shows the ATEs of Yelp's advertising package on purchase intention outcomes estimated using equation (1). Columns (1)–(4) show that advertising increases a restaurant's Yelp page views by 0.09 standard deviations (19%), map inquiries by 0.06 standard deviations (14%), calls by 0.03 standard deviations (7%), and clicks on restaurants' own websites by 0.03 standard deviations (7%).[13] We also find that advertising increases the monthly number of reviews left for the restaurant by 0.03 standard deviations (5%), suggesting that more customers ultimately visited the restaurant due to advertising.[14] Effects are statistically significant for all measures except for orders and reservations, for which we are underpowered to detect small effects.[15]

We observe limited substitution of paid page views for organic page views, suggesting that

[13]The effects on page views and leads allow us to see if paid page views convert customers down the funnel as effectively as organic page views (The conceptualization and calculation are detailed in Appendix B.) The impacts of search ads on conversion rates are shown in Appendix C Table C1 and Figure C1 Panel B. The smaller effects on leads compared with page views are associated with a 12% decrease in the overall conversion rate from page views to leads due to advertising, reflecting a 46% loss in this conversion rate for paid page views compared with organic page views. (Note: The 46% loss is the percentage relative to the baseline organic conversion rate, not the absolute loss.) This could be driven by consumers clicking on the ads by mistake or by worse matches between consumers and restaurants through ads, which is consistent with the findings in Johnson et al. (2017a) that incremental visitors from display ads are less likely to convert than baseline visitors.

[14]While advertising leads to more reviews, we find no statistically significant impacts on review ratings (see Appendix C Table C3), suggesting that customers who eventually visit the restaurant due to ads are not worse matches than customers who find the restaurant through organic search.

[15]Given the sample size of 5,867 restaurants for which we have data on orders, the minimum detectable effect at a significance level of 0.05 and 80% power is approximately 0.046 standard deviations. For reservations, for which we have data on 306 restaurants, the minimum detectable effect is approximately 0.2 standard deviations. In Appendix G Table G1, we also show subsample ATE on page views and leads. The effect sizes are similar for EAT24 and Yelp Reservations restaurants, so we do not believe that the insignificant effects on take-out orders and reservations are due to differences in the type of businesses.

advertising increased demand from customers beyond those that restaurants would have attracted without ads (Appendix C Table C2 and Figure C1 Panel A).[16] We also find that these treatment effects disappear immediately after the experiment ends, as shown in Figure 2 and Table 1. However, these patterns are not sufficient to reject the existence of the carryover effects of advertising: if converted consumers visit the restaurant after the experiment without visiting its Yelp page, our measures would not capture these consumers.

### 4.2 Heterogeneous Advertising Effects

In this section, we explore the heterogeneity in advertising effects along the following dimensions: restaurant age, chain status, review rating, the number of reviews, organic traffic levels, and price levels.

We construct variables on business attributes using data from the Yelp platform immediately before the start of the experiment. The restaurant's review rating is a single aggregate rating posted on the restaurant's Yelp page (a simple average of all past review ratings), and the number of reviews counts the total number of reviews on the restaurant. Restaurant age is calculated as the number of years the restaurant has operated. Given that Yelp started recording new restaurant entries in 2005 when it was founded, we use a dummy variable for restaurants that entered the market before 2005 to indicate that they have operated for more than ten years. To determine the chain status of each restaurant, we first flagged businesses that had three or more establishments with the same name in a city. We then checked each restaurant's website to determine whether the restaurant belonged to a chain and whether the chain operated locally or nationally.[17] Restaurant price level is the number of dollar signs displayed on Yelp, which ranges from $ (least expensive, or less than $10 per meal) to $$$$ (most expensive, or greater than $60 per meal). We also measure each restaurant's total organic page views during the three months before the experiment (between May 2015 and July 2015), and refer to this as pre-experiment organic traffic levels. Unlike all the other attributes, organic traffic is not directly observable to consumers. Appendix A Figure A1 and Table A3 show the distribution of these restaurant attributes.

We first examine heterogeneous advertising effects on purchase intention measures based on each restaurant attribute separately, which are reported in Table 2, Table 3, and Table 4. Qualita-

[16] At the time of the experiment, Yelp did not place a business in a paid search ad if it already appeared in the top three positions in an organic search ad, reducing such substitution.

[17] We call restaurants that operate in multiple locations within a region "local chains", and chain restaurants that have a national presence "national chains."

tively, we find larger effects on page views and leads for restaurants that are not national chains, higher-rated, more reviewed, newer, higher-priced, as well as those that have higher pre-experiment organic traffic. We test the statistical significance of these heterogeneous advertising effects and control the false discovery rate of multiple testing to be under 0.05 using the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995). Specifically, we conduct pairwise comparisons between all effects estimated in Table 2-4. Overall, we find that more than half of the pairwise differences are statistically significant when we compare restaurants with different chain statuses, ratings, number of reviews, prices, and baseline organic traffic (Appendix H Table H1-H6). The adjustment method and p-values are detailed in Appendix H.

Comparing advertising gains across different attributes qualitatively, we find that the differences in gains associated with pre-experiment organic traffic are the largest. For example, advertising increased page views by 0.054 standard deviations and leads by 0.022 standard deviations for restaurants with organic traffic below the 25th percentile, but increased page views by 0.133 standard deviations and leads by 0.102 standard deviations for restaurants with organic traffic above the 90th percentile — increasing gains almost five-fold (Table 3). We test that the effects on page views and leads are greater for restaurants with the higher traffic and obtain p-values of 0.0001 and 0.0005 respectively (Appendix H Table H3), allowing us to reject both null hypotheses after adjusting for multiple testing. The number of reviews has the second largest impact on advertising gains (Table 3). Other attributes differ in their importance in affecting advertising gains in page views versus leads qualitatively. For example, chain status affects advertising gains in page views more than in leads, whereas the restaurant's review rating affects advertising gains in leads more than in page views (Table 2).[18] Lastly, advertising gains also differ depending on the restaurant's age and price, but the magnitudes of differences are smaller than other dimensions, and the differences based on age are mostly not statistically significant.

We also present exploratory evidence on interaction effects between restaurant attributes to the extent that we are able with our data.[19] We first explore whether gains from the restaurant's pre-experiment traffic on advertising gains are driven by restaurants with higher review ratings

[18] 99.3% of the restaurants in the experiment have posted review ratings of 3 stars or more due to Yelp's rating requirement for the advertising businesses. Including all restaurants below 3 stars (all these are 2.5-star restaurants) in the group of 3-star restaurants had very little impact on the results.

[19] There is a limitation to examining heterogeneous effects based on the interaction of attributes. The more dimensions we include in the interaction to analyze heterogeneous effects, the fewer businesses we are left with within each cell of combined attributes, and the noisier the estimates become. In the results in appendix E, we at most interact three attributes at a time. Because of the reduced statistical power, the comparisons between effect sizes are qualitative.

(Appendix E). Appendix E Table E1 shows that conditional on the same traffic level, a higher rating often leads to greater advertising gains.[20] Conditional on the rating, higher traffic contributes to higher gains for restaurants with ratings of 3.5 stars or higher but not for restaurants with 3 stars, which is below the median rating on Yelp.

We also explore how advertising gains for chain and non-chain businesses vary. Large national chain restaurants have standardized menus and shared reputation and brand value (Klopack, 2018), and the literature has found that the returns to online reputation are lower for chains than independent businesses (Luca, 2018; Hollenbeck, 2018). Motivated by these findings, we explore how the gains for chain businesses vary with their review ratings and age. In Appendix E Figure E2, we plot advertising gains in page views and leads by chain status (demarcated by the x-axis position) and ratings (demarcated by color). The figure shows that ratings positively affect advertising gains for independent restaurants but provide limited gains for national chains. The results are robust to conditioning on the number of reviews (Appendix E Table E4). Similarly, we examine the effect of age on advertising gains by chain status. Whether we further condition on the number of reviews (Appendix E Table E5) or review ratings (Appendix E Table E6), we only find higher advertising gains for newer restaurants (who entered the market within two years of the experiment) if the restaurant does not belong to a national chain, consistent with the interpretation that the information value of ads may be lower for chains.

While our heterogeneous effects are exploratory, the findings can be put into the context of theories of search advertising. For instance, Lovett and Staelin (2016) and Sahni and Nair (2020) point to three roles of advertising—signaling, informing, and reminding. We find evidence consistent with all three roles of advertising. Both the informing and signaling effects work through updating the consumer's expectation about the advertiser's quality, which would suggest that both effects are stronger when there is greater uncertainty about the advertiser's quality. Consistent with this interpretation, we find that independent restaurants gain more from advertising. However, we also find greater advertising effects for restaurants with more reviews and higher pre-experiment traffic, who are likely to have higher quality and lower quality uncertainty. One possibility is that this results from a reminding effect. By making the business salient to consumers, the restaurant may attract consumers who already know its high quality and are trying to decide which restaurant to go to,[21] or attract consumers who do not know its quality by providing useful information (such as

[20] The advertising gain stops increasing beyond 4 stars (Table 2), so we grouped all restaurants with 4 or more stars into one level.

[21] Since the choice set of restaurants is often large, even in a local market, it could be hard for any single restaurant

reviews and photos) that allow consumers to learn about its high quality.

## 5 Discussion

We evaluate heterogeneity in the impact of digital advertising through a field experiment across over 18,000 restaurants on the Yelp platform. Our findings highlight that paid search advertising may be more effective for independent firms and those with higher-quality.

We conduct some back-of-the-envelope calculations, which provide some suggestive estimates that the gains in leads that we identify may map to a return on investment (ROI) of around 88.2%, a sizeable return (Appendix D). However, a key limitation of our study is that we do not observe all purchases at the end of the funnel and thus cannot estimate actual returns. Furthermore, our results do not provide insight into the long-term returns to advertising, which would understate the overall effect of ads. This suggests that further research exploring heterogeneity in the returns to advertising over time, ideally with granular data across the funnel, would provide valuable insights.

Lastly, our findings have practical implications for platforms that rely on advertising as a main source of revenue, suggesting which types of businesses are likely to gain from advertisements. It also highlights ways for a platform to identify types of businesses that might benefit from increasing their advertising on the platform, which can be useful for platforms that are looking to identify potential advertisers. More broadly, our results propose that advertising may help independent and high quality businesses grow—potentially helping increase competition in the marketplace. Our findings also highlight that how ads are designed—both in terms of the rules for selecting when businesses can be featured in ads as well as what information the ad displays—may significantly affect which firms gain from advertising and grow, and may provide a promising direction for future work.


## Funding and Competing Interests

All the authors have consulted for tech companies, including Yelp, but their compensation and ability to publish did not depend on the results of this paper. Yelp provided engineering support and data for this experiment.


to always show up at the top of every consumer's choice set or every organic search result, so the advertising's reminding effect that prominently places a business into a user's consideration set can work well.

## References


Athey, S. and G. Imbens (2016). The econometrics of randomized experiments. *arXiv preprint arXiv:1607.00698*.

Bagwell, K. (2007). The economic analysis of advertising. Volume 3, Chapter 28, pp. 1701–1844. Elsevier.

Benjamini, Y. and Y. Hochberg (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological) 57*(1), 289–300.

Benjamini, Y. and D. Yekutieli (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of statistics*, 1165–1188.

Biswas, S. (2020). *Investigating the effects of including discount information in advertising*. Ph. D. thesis, The University of Chicago.

Blake, T., C. Nosko, and S. Tadelis (2015). Consumer heterogeneity and paid search effectiveness: A large-scale field experiment. *Econometrica 83*(1), 155–174.

Campbell, C., J. Runge, K. Bates, S. Haefele, and N. Jayaraman (2021). It's time to close the experimentation gap in advertising: Confronting myths surrounding ad testing. *Business Horizons*.

Chesnes, M., W. D. Dai, and G. Jin (2017). Banning foreign pharmacies from sponsored search: The online consumer response. *Marketing Science 36*(6), 879–907.

Choi, H., C. F. Mela, S. R. Balseiro, and A. Leary (2020). Online display advertising markets: A literature review and future directions. *Information Systems Research 31*(2), 556–575.

Coviello, L., U. Gneezy, and L. Goette (2017). A large-scale field experiment to evaluate the effectiveness of paid search advertising.

Feit, E. M. and R. Berman (2019). Test & roll: Profit-maximizing a/b tests. *Marketing Science 38*(6), 1038–1058.

Glaeser, E. L., H. Kim, and M. Luca (2022). Nowcasting the local economy: Using yelp data to measure economic activity. *NBER/CRIW Volume for Big Data for 21st Century Statistics*.

Goldfarb, A. and C. Tucker (2011). Online display advertising: Targeting and obtrusiveness. *Marketing Science 30*(3), 389–404.

Gordon, B. R., K. Jerath, Z. Katona, S. Narayanan, J. Shin, and K. C. Wilbur (2021). Inefficiencies in digital advertising markets. *Journal of Marketing 85*(1), 7–25.

Gordon, B. R., R. Moakler, and F. Zettelmeyer (2022). Close enough? a large-scale exploration of non-experimental approaches to advertising measurement. *arXiv preprint arXiv:2201.07055*.

Gordon, B. R., F. Zettelmeyer, N. Bhargava, and D. Chapsky (2019). A comparison of approaches to advertising measurement: Evidence from big field experiments at facebook. *Marketing Science 38*(2), 193–225.

Graepel, T., J. Q. n. Candela, T. Borchert, and R. Herbrich (2010). Web-scale bayesian click-through rate prediction for sponsored search advertising in microsoft's bing search engine. In *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ICML'10, USA, pp. 13–20. Omnipress.

Ha, A. (2019). Us digital advertising exceeded $100b in 2018 (iab report). *TechCrunch*.

Hollenbeck, B. (2018). Online reputation mechanisms and the decreasing value of chain affiliation. *Journal of Marketing Research 55*(5), 636–654.

Hollenbeck, B., S. Moorthy, and D. Proserpio (2019). Advertising strategy in the presence of reviews: An empirical analysis. *Marketing Science 38*(5), 793–811.

Imbens, G. W. and D. B. Rubin (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Jeziorski, P. and S. Moorthy (2017). Advertiser prominence effects in search advertising. Technical report.

Johnson, G., R. A. Lewis, and E. Nubbemeyer (2017a). The online display ad effectiveness funnel & carryover: Lessons from 432 field experiments. *Available at SSRN 2701578*.

Johnson, G., S. Shriver, and S. Goldberg (2022). Privacy & market concentration: Intended & unintended consequences of the gdpr. *Available at SSRN 3477686*.

Johnson, G. A., R. A. Lewis, and E. I. Nubbemeyer (2017b). Ghost ads: Improving the economics of measuring online ad effectiveness. *Journal of Marketing Research 54*(6), 867–884.

Johnson, G. A., R. A. Lewis, and D. Reiley (2014). Location, location, location: repetition and proximity increase advertising effectiveness. *Available at SSRN 2268215*.

Johnson, G. A., R. A. Lewis, and D. H. Reiley (2017). When less is more: Data and power in advertising experiments. *Marketing Science 36*(1), 43–53.

Kalyanam, K., J. McAteer, J. Marek, J. Hodges, and L. Lin (2018). Cross channel effects of search engine advertising on brick & mortar retail sales: Meta analysis of large scale field experiments on google. com. *Quantitative Marketing and Economics 16*(1), 1–42.

Klopack, B. (2018). One size fits all? the value of standardized retail chains. Technical report, working paper.

Kohavi, R., D. Tang, and Y. Xu (2020). *Trustworthy online controlled experiments: A practical guide to a/b testing*. Cambridge University Press.

Lewis, R. A. and J. M. Rao (2015). The unfavorable economics of measuring the returns to advertising. *The Quarterly Journal of Economics*, qjv023.

Lewis, R. A., J. M. Rao, and D. H. Reiley (2011). Here, there, and everywhere: Correlated online behaviors can lead to overestimates of the effects of advertising. In *Proceedings of the 20th International Conference on World Wide Web*, WWW '11, New York, NY, USA, pp. 157–166. ACM.

Lewis, R. A. and D. H. Reiley (2014). Online ads and offline sales: measuring the effect of retail advertising via a controlled experiment on yahoo! *Quantitative Marketing and Economics 12*(3), 235–266.

Lovett, M. J. and R. Staelin (2016). The role of paid, earned, and owned media in building entertainment brands: Reminding, informing, and enhancing enjoyment. *Marketing Science 35*(1), 142–157.

Luca, M. (2018). Reviews, reputation, and revenue: The case of yelp.com. *HBS Working Paper*.

Luca, M. and G. Zervas (2016). Fake it till you make it: Reputation, competition, and yelp review fraud. *Management Science 62*(12), 3412–3427.

Mayzlin, D., Y. Dover, and J. Chevalier (2014). Promotional reviews: An empirical investigation of online review manipulation. *American Economic Review 104*(8), 2421–55.

McMahan, H. B., G. Holt, D. Sculley, M. Young, D. Ebner, J. Grady, L. Nie, T. Phillips, E. Davydov, D. Golovin, S. Chikkerur, D. Liu, M. Wattenberg, A. M. Hrafnkelsson, T. Boulos, and J. Kubica (2013). Ad click prediction: A view from the trenches. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '13, New York, NY, USA, pp. 1222–1230. ACM.

Nelson, L., D. Simester, and K. Sudhir (2020). Introduction to the special issue on marketing science and field experiments. *Marketing Science 39*(6), 1033–1038.

Nelson, P. (1974). Advertising as information. *Journal of Political Economy 82*(4), 729–54.

Rao, J. M. and A. Simonov (2019). Firms' reactions to public information on business practices: The case of search advertising. *Quantitative Marketing and Economics 17*(2), 105–134.

Sahni, N. S. (2015). Effect of temporal spacing between advertising exposures: Evidence from online field experiments. *Quantitative Marketing and Economics 13*(3), 203–247.

Sahni, N. S. and H. S. Nair (2020). Does advertising serve as a signal? evidence from a field experiment in mobile search. *The Review of Economic Studies 87*(3), 1529–1564.

Sahni, N. S., S. Narayanan, and K. Kalyanam (2019). An experimental investigation of the effects of retargeted advertising: The role of frequency and timing. *Journal of Marketing Research 56*(3), 401–418.

Shapiro, B. T., G. J. Hitsch, and A. E. Tuchman (2021). Tv advertising effectiveness and profitability: Generalizable results from 288 brands. *Econometrica 89*(4), 1855–1879.

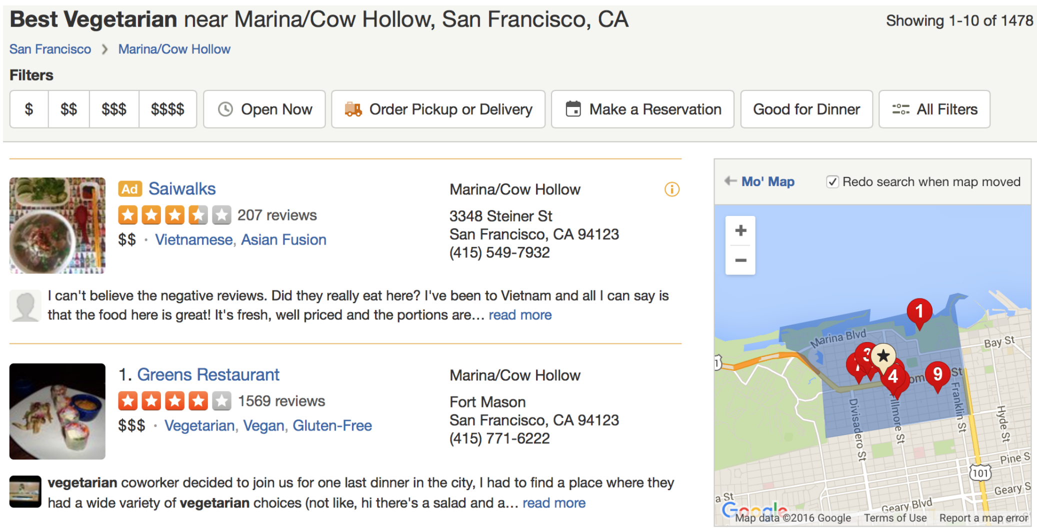
Simonov, A., C. Nosko, and J. M. Rao (2018). Competition and crowd-out for brand keywords in sponsored search. *Marketing Science 37*(2), 200–215.

Zervas, G., D. Proserpio, and J. W. Byers (2017). The rise of the sharing economy: Estimating the impact of airbnb on the hotel industry. *Journal of marketing research 54*(5), 687–705.

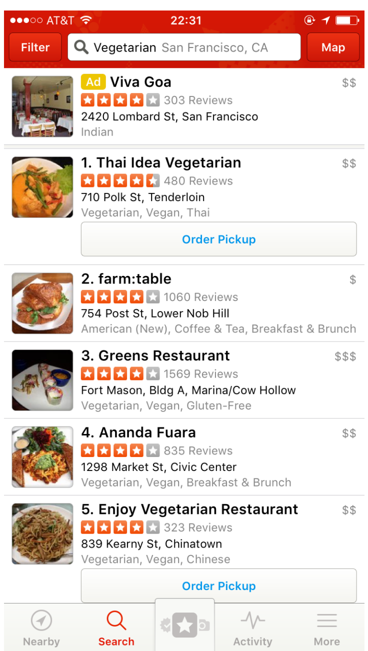
# Figures and Tables

Figure 1: Screenshots of Paid Search Ads on Yelp

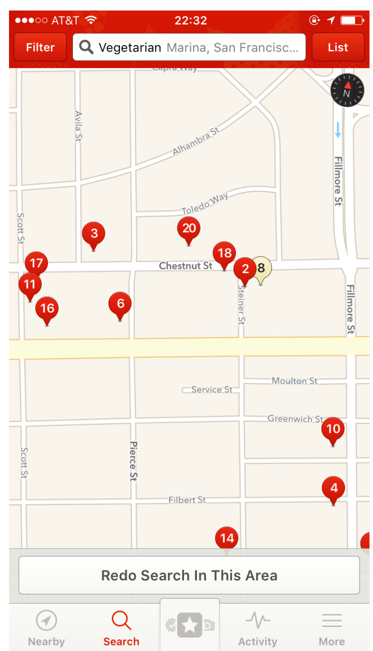
**A.** Paid Search Ad in Web Browser




**B.** Paid Search Ad in Mobile App



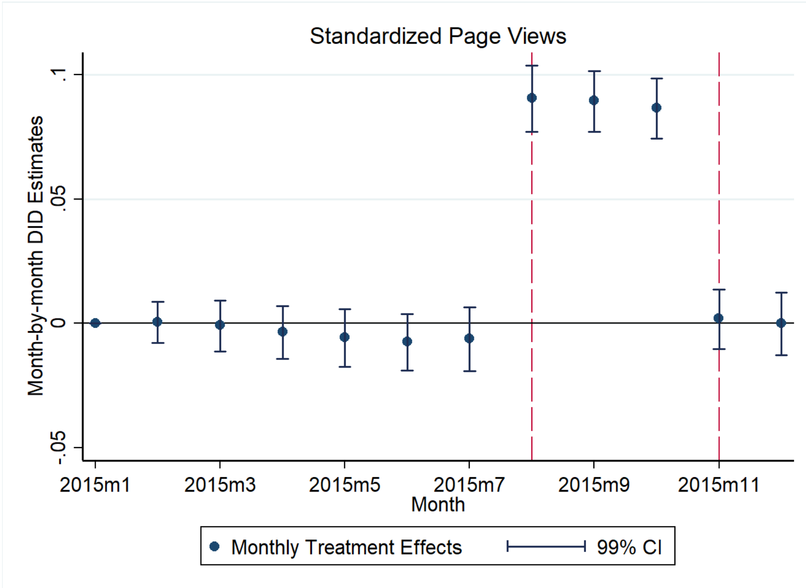

**C.** Paid Search Ad in Mobile App Map



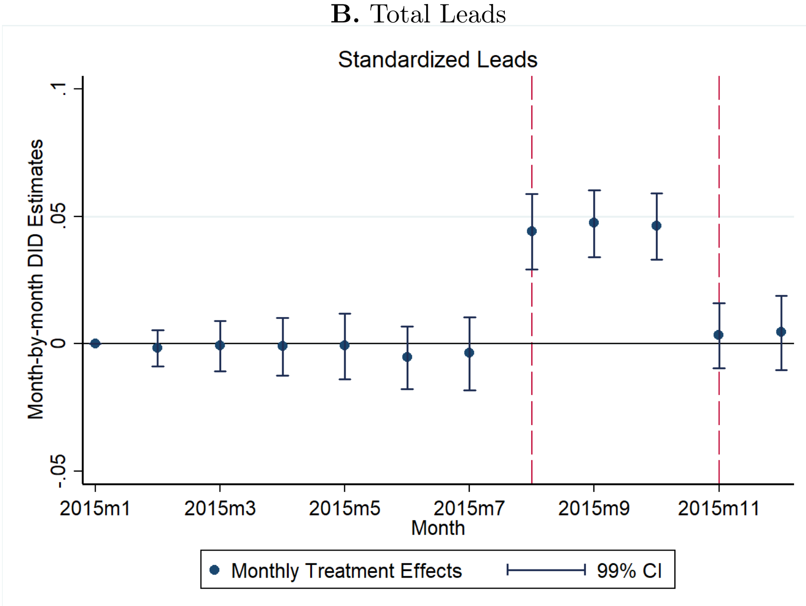

*Notes:* These figures show examples of result screens following the search for "vegetarian" near Marina, San Francisco, California. The advertised restaurant is shown in the paid search slot (the first link) on the page. A yellow flag "Ad" is shown next to the advertised restaurant's name that distinguishes the paid link from the organic links. During the period of our data collection, one slot, at most, is allocated to paid search ads in a search session.

Figure 2: Month-by-Month Treatment Effects on Standardized Page Views and Leads

**A.** Total Page Views




**B.** Total Leads




*Notes:* The figures plot treatment effects by month using difference-in-differences (specified in equation (1)), where the first month in the sample (January 2015) serves as the baseline. Each outcome is standardized by its standard deviation before the experiment. The two dashed vertical lines indicate the beginning and end of the experiment. See Appendix G Figure G1 and G2 for plots of standardized total page views and leads by subsample.

Table 1: Average Treatment Effects on Consumer Demand Measures

*[Outcomes Standardized]*

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| *Weighted Regressors*[a] | Page Views | Map Views | Calls from Mobile | External URL Clicks | Reviews | Orders | Reservations |
| Treated×ExprOn | 0.092*** | 0.064*** | 0.031*** | 0.032*** | 0.027*** | 0.017 | -0.101 |
| | (0.003) | (0.003) | (0.003) | (0.004) | (0.005) | (0.018) | (0.078) |
| Treated×PostExpr | 0.004 | 0.008* | 0.004 | 0.003 | 0.002 | 0.023 | -0.144 |
| | (0.004) | (0.004) | (0.005) | (0.005) | (0.006) | (0.024) | (0.086) |
| Other controls & business FE, month FE | x | x | x | x | x | x | x |
| N | 218,821 | 218,821 | 218,821 | 218,821 | 218,821 | 70,294 | 3,620 |
| *Mean effects in percentages and their 95% confidence intervals* | | | | | | | |
| During experiment | 18.8% | 13.8% | 6.5% | 6.8% | 5.0% | 1.5% | -15.7% |
| | [17.7%,19.9%] | [12.35%,15.1%] | [5.2%,7.9%] | [4.5%,9.1%] | [3.1%,6.9%] | [-1.5%,4.7%] | [-40.3%,8.1%] |
| Post experiment | 0.1% | 2.0% | 1.1% | 0.3% | 1.1% | 1.9% | -22.7% |
| | [-0.5%,2.5%] | [0.01%,4.0%] | [-0.6%,2.8%] | [-2.2%,3.1%] | [-1.9%,3.7%] | [-2.1%,5.8%] | [-51.8%, 3.2%] |

* $p<0.05$, ** $p<0.01$, *** $p<0.001$. Standard errors are in parentheses, clustered at the business level.

[a] Regressors are weighted according to regression equation (1) in columns (1)–(5), so the coefficients of Treated×ExprOn and Treated×PostExpr estimate the ATEs during and after the experiment. Regressors in columns (6) and (7) are not weighted since they only use data from a single subsample.

*Notes:* **1** Columns (1)–(5) show the results of running regression equation (1) that estimates the ATEs on consumer demand measures. Columns (6) and (7) run the typical DID specification $y_{it} = \sum_{\iota=1}^{2} \alpha_\iota \omega_i T_{\iota t} + \mu_i + \delta_t + \epsilon_{it}$ on the subsamples of EAT24 restaurants (for take-out orders) and Yelp Reservations restaurants (for reservations). The observation is at the business×month level, and the outcomes are standardized by the standard deviation of the overall sample before the experiment (January to July 2015). **2** The bottom half of the table reports the effects of paid ads in percentage terms. The 95% confidence interval of the estimates is calculated using bootstrap. **3** The panel is strictly balanced between June and December 2015, but if a restaurant has joined Yelp between January and June 2015, which accounts for 1.5% of the restaurants in the sample, we keep them in the regression. For these restaurants, we do not have the full 12 months of observations.

Table 2: Heterogeneous Advertising Effects by Chain Status and Review Rating

**Panel** A. Heterogeneous Effects by Chain Status

*[Outcomes Standardized]*

| *Weighted Regressors*[a] | (1) Page Views | (2) Leads | (3) Reviews |
|---|---|---|---|
| ExprOn×Treated | | | |
| ×I(non-chain) | 0.095*** | 0.049*** | 0.027*** |
| | (0.003) | (0.003) | (0.006) |
| ×I(local chain) | 0.095*** | 0.053*** | 0.041* |
| | (0.007) | (0.008) | (0.018) |
| ×I(national chain) | 0.044*** | 0.023*** | 0.013 |
| | (0.004) | (0.004) | (0.011) |
| Other Controls, Fixed Effects | x | x | x |
| N | 218,821 | 218,821 | 218,821 |

**Panel B**. Heterogeneous Effects by Review Ratings

*[Outcomes Standardized]*

| *Weighted Regressors*[a] | (1) Page Views | (2) Leads | (3) Reviews |
|---|---|---|---|
| ExprOn×Treated | | | |
| ×I(≤3 Stars) | 0.071*** | 0.021*** | 0.009 |
| | (0.004) | (0.003) | (0.009) |
| ×I(3.5 Stars) | 0.092*** | 0.041*** | 0.023** |
| | (0.004) | (0.004) | (0.007) |
| ×I(4 Star) | 0.101*** | 0.062*** | 0.025** |
| | (0.005) | (0.006) | (0.009) |
| ×I(>4 Star) | 0.097*** | 0.062*** | 0.075*** |
| | (0.010) | (0.013) | (0.022) |
| Other Controls, Fixed Effects | x | x | x |
| N | 216,586 | 216,586 | 216,586 |

* $p<0.05$, ** $p<0.01$, *** $p<0.001$. Standard errors are in parentheses, clustered at the business level.

[a] Regressors are weighted according to regression equation (2). We fully interact ExprOn×Treated with the dummy variables representing all levels of an attribute, so the coefficients shown in the above table represent heterogeneous advertising effects for restaurants with each level of an attribute. The coefficients can be directly compared to each other in showing what type of restaurants gain more from advertising.

*Notes:* **1** Observations are at the business×month level. Column (1) - (3) show heterogeneous advertising effects on total page views, leads (the summation of map views, calls to the restaurant, and restaurant external URL clicks), and the number of new reviews. **2** Panel A shows advertising effects for restaurants with different chain statuses. Panel B shows advertising effects for restaurants with different average review ratings. Due to Yelp's requirement for advertising businesses, most restaurants have at least a 3-star average rating. In the bottom category of $I(\leq 3Stars)$, 95% of the restaurants have a 3-star average rating and 5% have a 2.5-star average rating.

Table 3: Heterogeneous Advertising Effects by Baseline Organic Traffic and Number of Reviews

**Panel A**. Heterogeneous Effects by Baseline Organic Traffic

*[Outcomes Standardized]*

| *Weighted Regressors*[a] | (1) Page Views | (2) Leads | (3) Reviews |
|---|---|---|---|
| ExprOn×Treated | | | |
| ×I(OrgnicTraffic≤25th pctl) | 0.054*** | 0.022*** | 0.016*** |
| | (0.001) | (0.001) | (0.004) |
| ×I(OrgnicTraffic∈(25th,50th] pctl) | 0.087*** | 0.034*** | 0.010 |
| | (0.002) | (0.002) | (0.006) |
| ×I(OrgnicTraffic∈(50th,75th] pctl) | 0.106*** | 0.052*** | 0.028*** |
| | (0.004) | (0.004) | (0.008) |
| ×I(OrgnicTraffic∈(75th,90th] pctl) | 0.107*** | 0.065*** | 0.041** |
| | (0.007) | (0.008) | (0.015) |
| ×I(OrgnicTraffic>90th) pctl) | 0.133*** | 0.102*** | 0.062 |
| | (0.021) | (0.024) | (0.036) |
| Other Controls, Fixed Effects | x | x | x |
| N | 218,821 | 218,821 | 218,821 |

**Panel B**. Heterogeneous Effects by Number of Reviews

*[Outcomes Standardized]*

| *Weighted Regressors*[a] | (1) Page Views | (2) Leads | (3) Reviews |
|---|---|---|---|
| ExprOn×Treated | | | |
| ×I(#Rev≤25th pctl) | 0.061*** | 0.027*** | 0.018*** |
| | (0.002) | (0.002) | (0.004) |
| ×I(#Rev∈(25th,50th]) pctl) | 0.089*** | 0.039*** | 0.023** |
| | (0.004) | (0.004) | (0.007) |
| ×I(#Rev∈(50th,75th] pctl) | 0.102*** | 0.049*** | 0.029** |
| | (0.005) | (0.006) | (0.009) |
| ×I(#Rev∈(75th,90th] pctl) | 0.114*** | 0.066*** | 0.036* |
| | (0.008) | (0.009) | (0.015) |
| ×I(#Rev>90th pctl) | 0.116*** | 0.086*** | 0.038 |
| | (0.017) | (0.019) | (0.033) |
| Other Controls, Fixed Effects | x | x | x |
| N | 218,821 | 218,821 | 218,821 |

* $p<0.05$, ** $p<0.01$, *** $p<0.001$. Standard errors are in parentheses, clustered at the business level.

[a] Regressors are weighted according to regression equation (2). We fully interact ExprOn×Treated with the dummy variables representing all levels of an attribute, so the coefficients shown in the above table represent heterogeneous advertising effects for restaurants with each level of an attribute. The coefficients can be directly compared to each other in showing what type of restaurants gain more from advertising.

*Notes:* **1** Observations are at the business×month level. Column (1) - (3) show heterogeneous advertising effects on total page views, leads (the summation of map views, calls to the restaurant, and restaurant external URL clicks), and the number of new reviews. **2** In Panel A, we split the sample based on the restaurant's rank by its 3-month pre-experiment organic page views. In panel B, we split the sample based on the restaurant's rank in terms of review numbers and show advertising effects for restaurants with different numbers of reviews.

Table 4: Heterogeneous Advertising Effects by Restaurant Age and Price Category

**Panel A**. Heterogeneous Effects by Restaurant Age

*[Outcomes Standardized]*

| *Weighted Regressors*[a] | (1) Page Views | (2) Leads | (3) Reviews |
|---|---|---|---|
| ExprOn×Treated | | | |
| ×I(age<1yr) | 0.119*** | 0.069** | 0.066 |
| | (0.023) | (0.024) | (0.037) |
| ×I(age∈[1,2]yr) | 0.103*** | 0.066*** | 0.053* |
| | (0.013) | (0.015) | (0.021) |
| ×I(age∈[3,4]yr) | 0.090*** | 0.032*** | 0.030* |
| | (0.007) | (0.007) | (0.013) |
| ×I(age∈[5,10]yr) | 0.097*** | 0.054*** | 0.029*** |
| | (0.004) | (0.004) | (0.008) |
| ×I(age>10yr) | 0.081*** | 0.040*** | 0.008 |
| | (0.003) | (0.003) | (0.007) |
| Other Controls, Fixed Effects | x | x | x |
| N | 218,266 | 218,266 | 218,266 |

**Panel** B. Heterogeneous Effects by Price Category

*[Outcomes Standardized]*

| *Weighted Regressors*[a] | (1) Page Views | (2) Leads | (3) Reviews |
|---|---|---|---|
| ExprOn×Treated | | | |
| ×I(missing) | 0.047*** | 0.019*** | 0.002 |
| | (0.011) | (0.006) | (0.009) |
| ×I($) | 0.078*** | 0.038*** | 0.024*** |
| | (0.003) | (0.003) | (0.006) |
| ×I($$) | 0.095*** | 0.051*** | 0.031*** |
| | (0.004) | (0.004) | (0.008) |
| ×I($$$) | 0.120*** | 0.064*** | 0.025 |
| | (0.011) | (0.012) | (0.017) |
| ×I($$$$) | 0.097*** | 0.022 | -0.005 |
| | (0.029) | (0.025) | (0.049) |
| Other Controls, Fixed Effects | x | x | x |
| N | 218,821 | 218,821 | 218,821 |

* $p<0.05$, ** $p<0.01$, *** $p<0.001$. Standard errors are in parentheses, clustered at the business level.

[a] Regressors are weighted according to regression equation (2). We fully interact ExprOn×Treated with the dummy variables representing all levels of an attribute, so the coefficients shown in the above table represent heterogeneous advertising effects for restaurants with each level of an attribute. The coefficients can be directly compared to each other in showing what type of restaurants gain more from advertising.

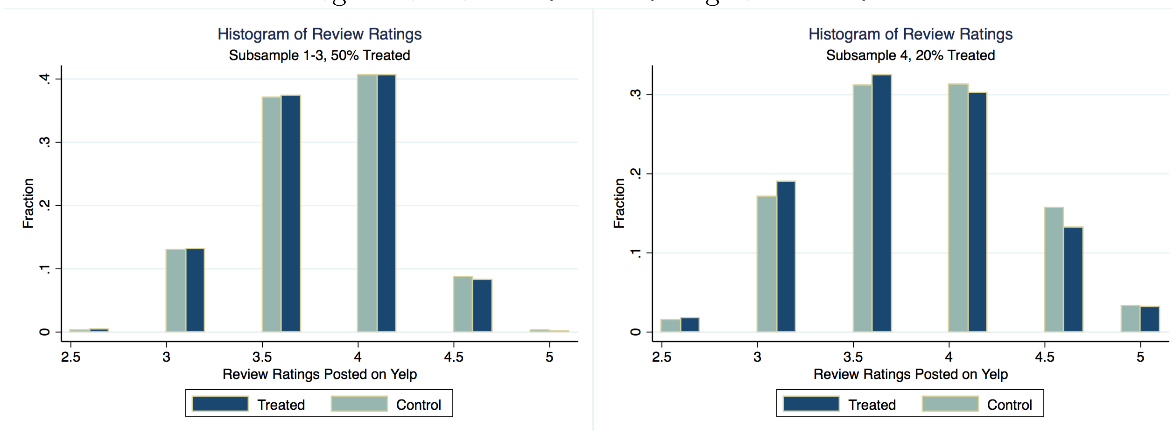
*Notes:* **1** Observations are at the business×month level. Column (1) - (3) show heterogeneous advertising effects on total page views, leads (the summation of map views, calls to the restaurant, and restaurant external URL clicks), and the number of new reviews. **2** Panel A and Panel B show advertising effects for restaurants with different ages and price categories respectively.

# Online Appendix: Which Firms Gain from Digital Advertising?

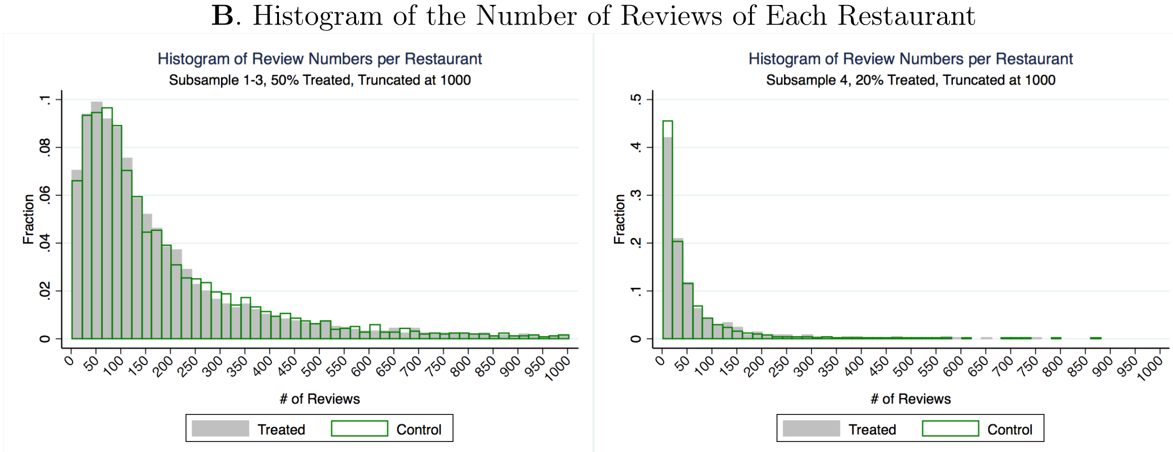
# Appendix A Experiment Design and Sample Details

Figure A1: Baseline Restaurant Review Attributes of Treated and Control Samples

**A.** Histogram of Posted Review Ratings of Each Restaurant




**B**. Histogram of the Number of Reviews of Each Restaurant




*Notes:* **1** Figure A plots the distribution of restaurant ratings before the experiment by treatment status. **2** Figure B plots the distribution of the number of reviews of the restaurants before the experiment by treatment status. Due to the sparse long right tail, the distribution of the number of reviews is truncated at 1,000. **3** Subsample 4 is plotted separately since its treated ratio is 20%, different from the 50% treated ratio in subsamples 1–3. With different treated ratios, comparing treated and control by simply pooling all subsamples together results in different subsample weights within the treated and control samples and invalidates the balance test. **4** The plots show the balance in restaurant attributes between the treated and control samples. The randomization test of outcome variables is shown in Appendix A Table A2.

Table A1: Subsamples in the Experiment

| Subsample | Description | Size | Treated | Control |
|---|---|---|---|---|
| 1. Yelp Reservations restaurants | Restaurants on Yelp that have partnered with Yelp's reservation service | 306 | 50% | 50% |
| 2. OpenTable restaurants | Restaurants on Yelp that have partnered with the OpenTable reservation service | 5,740 | 49% | 51% |
| 3. EAT24 restaurants | restaurants on Yelp that have partnered with Yelp's delivery service | 5,867 | 50% | 50% |
| 4. Non-partner restaurants in WA | Restaurants on Yelp that are non-partners and located in Washington state. | 6,381 | 20% | 80% |
| | Total: | 18,294 | 7,209 | 11,085 |

*Notes:* **1** Mutually exclusive subsamples of restaurants listed on Yelp are selected before the randomization. Restaurants are selected based on subsamples since certain outcomes of interests are available only in subsamples. For example, data on reservations are only available for restaurants that have partnered with the Yelp Reservations platform (subsample 1), and data on take-out orders are only available for restaurants that partnered with the Yelp EAT24 delivery platform (subsample 3). Historical restaurant revenue data are only available for restaurants in Washington state (WA) (subsample 4). **2** See Section 2 for details on how each subsample is constructed. **3** Within each subsample, treatment is randomly assigned to restaurants. Treated restaurants receive the Yelp business ad package for free for three months between August and October 2015.

Table A2: Randomization Check

| | $p$-value of $t$-test ($H_0 : \mu_{treated} = \mu_{control}$) | |
|---|---|---|
| | Partner Sample (Subsamples 1-3) | WA Non-Partner Sample (Subsample 4) |
| Outcome Variables | | |
| Page views | 0.401 | 0.496 |
| Map views | 0.398 | 0.695 |
| External URL clicks | 0.655 | 0.306 |
| Call from mobile | 0.712 | 0.541 |
| Number of reviews | 0.510 | 0.566 |
| Number of take-out orders | 0.947 | - |
| Number of reservations | 0.933 | - |
| Mean review rating | 0.130 | 0.452 |
| % of trusted reviews | 0.498 | 0.728 |
| Review length | 0.426 | 0.415 |

*Notes:* **1** The table reports $p$-values of tests of the null hypothesis that the mean of the treated and control sample outcomes are equal. **2** We show randomization test results for subsample 4 separately because of its different treated ratio. Pooling all subsamples together when treated ratios are different invalidates the balance test by putting different subsample weights across the treated and control samples. **3** Tests are conducted based on the sample outcomes in July 2015, the month before the experiment started. **4** Data on take-out orders are only available in subsample 3, and data on reservations are only available in subsample 1.

Table A3: Distribution of Restaurant Attributes

| Subsample | 1 | 2 | 3 | 4 | Full |
|---|---|---|---|---|---|
| Total | 306 | 5740 | 5867 | 6381 | 18294 |
| Chain status | | | | | |
| non-chain | 98% | 92% | 95% | 83% | 90% |
| local chain | 1% | 6% | 5% | 6% | 5% |
| national chain | 1% | 2% | 0% | 11% | 5% |
| Age (opens within) | | | | | |
| <1 year | 18% | 4% | 6% | 7% | 6% |
| 1-2 years | 15% | 9% | 11% | 8% | 9% |
| 3-4 years | 16% | 16% | 18% | 14% | 2% |
| 5-10 years | 38% | 48% | 40% | 34% | 41% |
| >10 years | 13% | 24% | 24% | 37% | 28% |
| Review rating | | | | | |
| <3.5 stars | 9% | 8% | 19% | 19% | 15% |
| 3.5 stars | 27% | 37% | 38% | 31% | 35% |
| 4 stars | 43% | 47% | 34% | 31% | 37% |
| >4 stars | 19% | 8% | 9% | 18% | 12% |
| missing | 0.7% | 0.1% | 0.0% | 0.6% | 0.3% |
| Number of reviews | | | | | |
| Tier 1 (in 0-25th percentile) | 13% | 1% | 18% | 54% | 25% |
| Tier 2 (in 25-50th percentile) | 22% | 15% | 32% | 28% | 25% |
| Tier 3 (in 50-75th percentile) | 23% | 33% | 30% | 12% | 25% |
| Tier 4 (in 75-100th percentile) | 41% | 51% | 20% | 5% | 25% |
| Pre-experiment Organic Traffic | | | | | |
| Tier 1 (in 0-25th percentile) | 10% | 1% | 15% | 57% | 25% |
| Tier 2 (in 25-50th percentile) | 14% | 15% | 34% | 26% | 25% |
| Tier 3 (in 50-75th percentile) | 20% | 35% | 29% | 12% | 25% |
| Tier 4 (in 75-100th percentile) | 56% | 49% | 22% | 5% | 25% |

*Notes:* The first four columns show, within each subsample, the fraction of restaurants that belongs to each different level of an attribute. The last column shows the fraction in the full sample. The fractions within each attribute dimension vertically add up to one. See Table A1 for the definition of different subsamples. On Yelp, the majority of restaurants are non-partners. On average, they have a lower online prominence than partner restaurants. As shown in the table above and in Panel B of Figure A1, restaurants in the fourth subsample have considerably fewer consumer reviews than restaurants in other subsamples.

## Appendix B. The Empirical Specification and Conversion Rate Calculation

### B.1 The Empirical Specification for Estimating the ATE and HTE

In this section, we explain the intuition underlying equation (1) that estimates the ATE. Equation (1) in Section 3 looks similar to the typical DID specification with two-way fixed effects. The only differences are the weights $\lambda_s$ in the terms interacting with the treatment status $\omega_i$. We briefly discuss the intuition of these weights here.

With stratified sampling, the overall sample mean of the treated restaurants weighs each subsample by the share of treated restaurants in each subsample. Similarly, the overall sample mean of the control restaurants weighs each subsample by the share of control restaurants in each subsample. If all subsamples have the same treated ratio, the usual DID specification works perfectly and the weights are equal to the sample size or share of each subsample. However, if the treated ratio differs by subsample, the weights assigned to calculate the sample mean for the treated and control samples are different, which leads to a biased estimator for the ATE. Hence, weights that are functions of subsample shares $\lambda_s$ must be added. Equation (1) is equivalent to first calculating the subsample ATEs and then aggregating them using the subsample shares as weights. Equation (1) can be easily verified by calculating and comparing the expected sample mean for each treated and control subsample during each period.

To estimate the HTEs, we deal with one dimension of restaurant attributes at a time and discretize each restaurant attribute to allow for non-monotonic HTEs. Following notations in Section 3, let the dummy variable $T_{\iota t}$, $\iota \in \{0, 1, 2\}$ denote whether month $t$ is in the period before, during, or after the experiment; the dummy variable $S_{si}$, $s \in \{1, 2, 3, 4\}$ denote the subsample that restaurant $i$ is in; and $\omega_i$, $\omega \in \{0, 1\}$ denote whether restaurant $i$ is assigned to the control or treatment group. In addition, let dummy variable $X_{zi}$, $z \in \{1, ..., Z\}$ denote categorical restaurant attributes. We then estimate HTEs for restaurants with different attributes using the following specification:

$$\begin{aligned} y_{it} =& \beta_0 + \sum_{\iota=1}^{2}\sum_{s=2}^{4}\gamma_{1s\iota}S_{si}T_{\iota t} + \sum_{\iota=1}^{2}\sum_{z=2}^{Z}\gamma_{2z\iota}X_{zi}T_{\iota t} + \sum_{\iota=1}^{2}\sum_{z=2}^{Z}\sum_{s=2}^{4}\gamma_{3sz\iota}S_{si}X_{zi}T_{\iota t} \\ &+ \sum_{z=1}^{Z}[\sum_{\iota=1}^{2}\alpha_{z\iota}\frac{1}{\phi_1}\omega_i S_{1i}X_{zi}T_{\iota t} + \sum_{\iota=1}^{2}\sum_{s=2}^{4}\beta_{sz\iota}(S_{si} - \frac{\phi_s}{\phi_1}S_{1i})\omega_i X_{zi}T_{\iota t}] + \alpha_i + \delta_t + \epsilon_{it}, \qquad (2) \end{aligned}$$

where

$$\phi_s = \frac{\lambda_s r_s^z}{\sum_{s=1}^{S} \lambda_s r_s^z}.$$

$r_s^z$ is the share of restaurants with attribute $z$ in subsample $s$ and $\lambda_s$ is the total subsample share, so $\phi_s$ is the share of restaurants with attribute $z$ in subsample $s$ among all restaurants with attribute $z$. In regression equation (2), the coefficients of interest are $\alpha_{1z}$, $\alpha_{2z}$ that estimate the HTE for restaurants with attribute $z$ during and after the experiment. In addition, $\beta_{s\iota z}$ estimates the HTE for restaurants with attribute $z$ in subsample $s$ ($s = 2, 3, 4$) during and after the experiment. In equation (2), the full interaction terms of dummy variables of $S_{si}$, $X_{zi}$, $T_{\iota t}$, and $\omega_i$ are included. Note that $\omega_i S_{si} T_{\iota t}$ and $\omega_i X_{zi} T_{\iota t}$ are absorbed since in the regressions with $\omega_i$ that estimate HTE, we include the full interactions of restaurant attributes $X_{zi}$ for $z = 1, 2, ..., Z$. So in the regression results, we show the mean HTE for every level of restaurant attributes rather than the comparisons against the baseline attributes.

## B.2 Calculating CTR and CTA Conversion Rates

We now back out how customers move along the paid search conversion funnel and compare it to the organic funnel. We focus the analysis on two conversion stages. The first conversion rate we measure is the CTR of paid links and organic links shown in the search result page. After the customer lands on the restaurant's page through an organic or paid click, she may interact with one of the CTA buttons by requesting the map direction to the restaurant, going to the restaurant's own URL, or calling the restaurant for more information. We call the conversion rate from page view to CTA engagements the CTA conversion rate. This is the second major conversion rate that we examine.

The following equation shows how search advertising affects leads. Let $\rho_{org}$ and $\rho_{paid}$ denote the CTA conversion rates of organic and paid page views, respectively. The generation of leads for a restaurant can then be written as

$$\begin{aligned} Leads &= PaidImpr \times PaidCTR \times \rho_{paid} + OrganicImpr \times OrganicCTR \times \rho_{org} \\ &= PaidPV \times \rho_{paid} + Traffic \times \rho_{org}. \end{aligned} \tag{3}$$

The equation shows that when search advertising is on, its effect on leads is larger if the restaurant gets a higher CTR of paid impressions and a higher conversion from page views to CTAs, and if the restaurant gets limited impact on the organic conversions. To see the mechanism of what types of restaurants benefit more from search ads, in what follows we unpack the impact of search advertising and zoom into the heterogeneous effect through each of these conversion channels based on restaurant attributes.

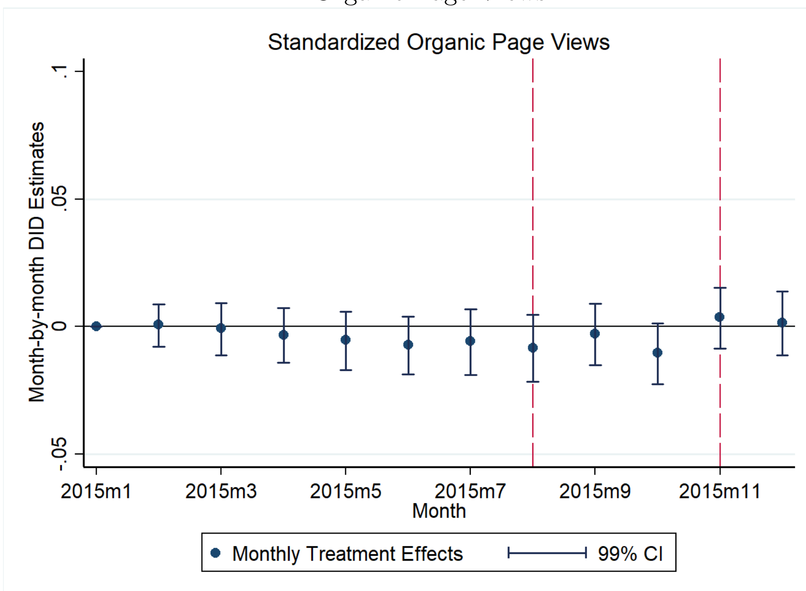
Below, we discuss the issues of calculating and analyzing the conversion rates. To calculate the CTRs of organic and paid links, in addition to page views, we collect data on paid and organic impressions—the number of times a restaurant appears in the paid or organic search result. A restaurant's paid or organic CTR is then calculated using the ratio of page views over impressions in a given time period. We want to calculate the CTA conversion rate similarly using the ratio of leads over page views, but we cannot calculate it for paid and organic page views separately due to a data limitation. We only observe the total number of leads for a restaurant's Yelp page and cannot trace back to what fraction of the leads are from organic or paid page views. To assess the ability for paid page views to convert into leads or demand, we have to make assumptions for this calculation.

The organic search team at Yelp works independently of the paid search team, and the placement of paid links does not directly affect organic search results. Unless paid search links divert otherwise uninterested consumers to click on the restaurant's organic link, we expect the behavior of consumers who are on the restaurant's Yelp page through organic link to remain the same. Hence, we assume that the CTA conversion rate of organic page views is unaffected by search advertising. With this assumption, we can easily calculate the CTA conversion rate of paid page views from the overall CTA conversion rate. If we divide equation (3) by the total number of page views and define $\rho_{all} = Leads/TotalPageViews$, we get $\rho_{all} = \eta_{paid} \times \rho_{paid} + (1 - \eta_{paid}) \times \rho_{org}$, where $\eta_{paid}$ is the fraction of page views coming from paid links. When we rearrange the equation, for the treated businesses with search advertising, we get $\rho_{paid} - \rho_{org} = \frac{\rho_{all} - \rho_{org}}{\eta_{paid}}$. The left-hand side of this equation gives us what we want to measure—how much the CTA conversion rate will be different for consumers who visit the page through a paid link instead of an organic link. With the experiment and the assumption that search advertising does not affect the CTA conversion rate of organic page views, $\rho_{org}$ of the treated restaurants can be approximated using $\rho_{org}$ of the control restaurants. Since control restaurants do not have paid search traffic and all page views come from organic links, $\rho_{org}$ is the same as $\rho_{all}$ for the control restaurants. In addition, the numerator of this equation is the treatment effect of search advertising on the overall CTA conversion rate.

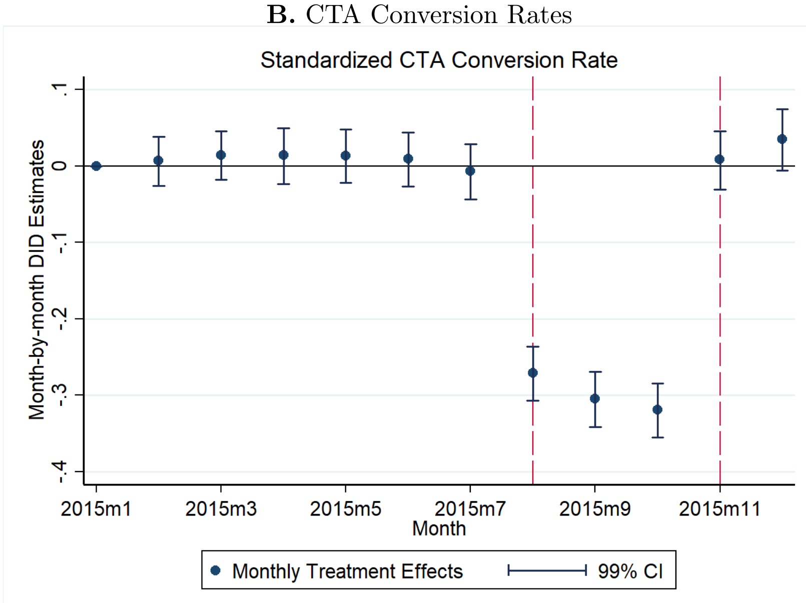
# Appendix C Additional Results on Average Treatment Effects

Figure C1: Month-by-Month Treatment Effects on Standardized Organic Page Views and CTA Conversion Rates

**A.** Organic Page Views




**B.** CTA Conversion Rates




*Notes:* **1** The figures plot the differential changes in the outcome of treated and control samples by month estimated by difference-in-differences, where the first month (January 2015) serves as the baseline. More specifically, we run the regression specified in equation (1) and estimate the treatment effect for each month. **2** Each outcome is standardized by its standard deviation before the experiment. **3** The two dashed vertical lines indicate the beginning and end of the experiment or the three-month advertising treatment. **4** See Appendix G Figure G3 for plots of standardized CTA conversion rates by subsample.

Table C1: Average Treatment Effects on Call-to-Action Conversion Rates

| *[Outcomes Standardized]* | | |
|---|---|---|
| *Weighted Regressors*[a] | Overall CTA Conversion Rate | |
| Treated×ExprOn | -0.306*** | |
| | (0.009) | |
| Treated×PostExpr | 0.014 | |
| | (0.011) | |
| Other controls & business FE, month FE | x | |
| N | 218,670 | |
| *Effects in Percentages* | | |
| | % Effects on Overall CTA Conversion Rate | % Change in CTA Conversion Rate (Paid vs. Organic)[b] |
| During experiment | -11.7% | -45.8% |
| | [-12.4%,-11.0%] | [-48.3%,-43.2%] |
| Post experiment | 0.06% | 2.1% |
| | [-0.3%,1.3%] | [-0.9%,4.8%] |

* $p<0.05$, ** $p<0.01$, *** $p<0.001$. Standard errors are in parentheses, clustered at the business level.

[a] Regressors are weighted according to regression equation (1). The coefficients of Treated×ExprOn and Treated×PostExpr estimates the ATEs during and after the experiment.

[b] $\rho_{org}$ and $\rho_{paid}$ denote the call-to-action (CTA) conversion rate of organic and paid page views, respectively. The percentage change in the CTA conversion rate (paid vs. organic) measures $(\rho_{paid}-\rho_{org})/\rho_{org}$ during search advertising and is calculated following the discussion in Appendix B.2.

*Notes:* **1** The upper panel shows the result of running regression equation (1) that estimates the ATE on the overall CTA conversion rate. The observation is at the business×month level, and the outcomes are standardized by the standard deviation of the overall sample before the experiment (January to July 2015). **2** The bottom half of the table reports the mean effects of paid search ads in percentage terms. The 95% confidence interval of the estimates is calculated using bootstrap.

**3** The effects by subsample are shown in Appendix G Table G2.

Table C2: Average Treatment Effects on Organic Impressions and Page Views

*[Outcomes Standardized]*

| *Weighted Regressors*[a] | (1) Organic Impression | (2) Organic page views | (3) Organic CTR |
|---|---|---|---|
| Treated×ExprOn | 0.0006 | -0.004 | -0.091* |
| | (0.0027) | (0.003) | (0.041) |
| Treated×PostExpr | -0.0014 | 0.006 | -0.157 |
| | (0.0046) | (0.004) | (0.085) |
| Other controls & business FE, month FE | x | x | x |
| N | 218,821 | 218,821 | 218,627 |

* $p<0.05$, ** $p<0.01$, *** $p<0.001$. Standard errors are in parentheses, clustered at the business level.

[a]Regressors are weighted according to regression equation (1), so the coefficients of Treated×ExprOn and Treated×PostExpr estimate the average treatment effects during and after the experiment.

*Notes:* The table shows the results of running regression equation (1) that estimates the average treatment effects. The observation is at the business×month level, and the outcomes are standardized by the standard deviation of the overall sample before the experiment (January to July 2015).

Table C3: Average Treatment Effect on New Reviews Written on the Restaurants

*[Outcomes Standardized]*

| *Weighted Regressors*[a] | (1) New Review Trustworthy% | (2) New Review Rating | (3) New Review Length |
|---|---|---|---|
| Treated×ExprOn | -0.013 | -0.007 | -0.013 |
| | (0.014) | (0.013) | (0.013) |
| Treated×PostExpr | 0.011 | 0.004 | 0.022 |
| | (0.017) | (0.017 ) | (0.016) |
| Other controls & business FE, month FE | x | x | x |
| N | 169,958 | 163,010 | 169,958 |

* $p<0.05$, ** $p<0.01$, *** $p<0.001$. Standard errors are in parentheses, clustered at the business level.

[a]Regressors are weighted according to regression equation (1), so the coefficients of Treated×ExprOn and Treated×PostExpr estimate the average treatment effects during and after the experiment.

*Notes:* The table shows the results of running the regression equation (1) that estimates the average treatment effects on new reviews written on the restaurants during the study period. The variable Trustworthy% is the percentage of new reviews labeled as trustworthy by Yelp. The observation is at the business×month level, and the outcomes are standardized by the standard deviation of the overall sample before the experiment (January to July 2015). The sample size in column (2) is smaller than in the other two regressions since some reviews are posted with the review rating missing.

Table C4: Average Treatment Effects Estimated Using Sample During the Experiment Period

*[Outcomes Standardized*

| *Weighted Regressors*$^a$ | (1) Page Views | (2) Map Views | (3) Calls from Mobile | (4) External URL Clicks | (5) Reviews |
|---|---|---|---|---|---|
| Treated | 0.091*** | 0.059*** | 0.031* | 0.036** | 0.028* |
| | (0.013) | (0.014) | (0.014) | (0.013) | (0.013) |
| Other Controls Month FE | x | x | x | x | x |
| N | 54,882 | 54,882 | 54,882 | 54,882 | 54,882 |

* p<0.05, ** p<0.01, *** p<0.001. Standard errors are in parentheses, clustered at the business level.
[a] In this table, we use the business-by-month level data only during the experiment period (3 months between August 2015 and October 2015) and we run the following regression:

$$y_{it} = \beta_0 + \alpha \frac{1}{\lambda_1} \omega_i S_{1i} + \sum_{s=2}^{4} \beta_s (S_{si} - \frac{\lambda_s}{\lambda_1} S_{1i}) \omega_i + \delta_t + \epsilon_{it},$$

where $S_s$, $s \in \{1, 2, 3, 4\}$ are dummy variables denoting each subsample, $i$ denotes a restaurant, $t$ denotes a month, $\omega$ denotes the treatment status, and $\lambda_s$ is the percentage experiment sample in subsample $s$. We added $\delta_t$ as month fixed effects to control for common time trend. In this equation, $\alpha$ estimates the average treatment effects during the experiment and report these estimates in the table above. This estimates from specification are equivalent to first taking the simple difference in outcomes between the treated and control groups within each subsample and then taking the weighted sum of these differences using subsample sizes as weights. The outcomes in this table are standardized by their pre-experiemnt standard deviation.

Table C5: Average Treatment Effects (Collapsing Pre-experiment, During-experiment, and Post-experiment Observations)

*[Outcomes Standardized*

| *Weighted Regressors*$^a$ | (1) Page Views | (2) Map Views | (3) Calls from Mobile | (4) External URL Clicks | (5) Reviews |
|---|---|---|---|---|---|
| Treated×ExprOn | 0.092*** | 0.064*** | 0.031*** | 0.032*** | 0.027*** |
| | (0.003) | (0.003) | (0.003) | (0.004) | (0.005) |
| Treated×PostExpr | 0.005 | 0.009* | 0.004 | 0.004 | 0.003 |
| | (0.004) | (0.004) | (0.006) | (0.005) | (0.007) |
| Other Controls, Period and Biz FE | x | x | x | x | x |
| N | 54,882 | 54,882 | 54,882 | 54,882 | 54,882 |

* p<0.05, ** p<0.01, *** p<0.001. Standard errors are in parentheses, clustered at the business level.
[a] In this table, we collapse the monthly data into three periods, pre-experiment, during-experiment, and post-experiment for each business. We then run similar regressions as the ones reported in Table 1 of the draft. Dummy variables indicating the during-experiment and post-experiment periods and business fixed effects are added.

## Appendix D Estimating Sales Returns to Advertising

To estimate sales returns to advertising, we conduct a back-of-the-envelope calculation by putting dollar values on the marginal increase in page views. We have historical tax revenue for the first two quarters of 2015 from the Washington State Department of Revenue (pre-experiment), which we are able to match with 835 Washington state restaurants in our sample (13% of subsample 4).[22]

Using this historical data, we estimate the relationship between sales and page views, which we use to approximate the change in revenue when page views increase. Specifically, we conduct the following regression to correlate revenue with page views:

$$Ln(Revenue_{it}) = \beta Ln(Pageview_{it}) + \alpha_i + \delta_t + \epsilon_{it}$$

for each restaurant $i$ in quarter $t$. We add restaurant fixed effects, so $\beta$ captures the revenue change related to changes in the restaurant's Yelp page views within a restaurant. We also add quarterly dummy variables to control for seasonality. The estimate for $\beta$ is 32.54%, statistically significant at the 1% level with standard errors clustered at the business level.

This exercise has a number of limitations. First, the match rate is very low at at 13%, due to the fact that the Washington state revenue data is recorded with the owner's name and address, which differ from the restaurant's name and address. This raises the question of selection bias in the matched sample, which we cannot assess empirically. Second, the estimate for $\beta$ is biased, given that a change in page views may be correlated with various restaurant attributes that can also lead to changes in revenue, and the fact that there may be reverse causality with a change in revenues leading to a change in page views. Third, this data is from the first two quarters of 2015, which is prior to our experiment that ran between August and October of 2015. The data therefore may not be representative of the correlation between changes in page views and revenues in the period during or following the experiment.

We thus use this exercise as a back-of-the-envelope calculation rather than as an accurate or precise estimate of the returns to advertising. In this light, the estimate suggests that a 10% increase in quarterly page views is correlated with a 3.3% increase in quarterly revenue. Since we find an ATE of an 18.8% increase in total page views (Table 1), this effect would suggest an increase of roughly 6.2% increase in revenue, which translates to a fairly large return to advertising. An

[22] The match rate is low since the name and address recorded in the revenue data are often the owner's business name and address, different from the restaurant's actual name and address.

important caveat is that marginal page views generated from search ads have lower conversion rates compared to organic page views. Thus, we adjust the revenue effect by applying the 45.8% loss in the conversion rate of paid page views, resulting in 3.36% increase in revenue. For example, for a business with $96,000 in quarterly sales (the median in Washington state), assuming the profit margin for additional sales is around 70%,[23] the gain from advertising is ($\$96,000 \times 3.36\% \times 70\% =)\$2,258$ for the quarter. The cost of advertising on Yelp is $1,200 per quarter, resulting in a return on investment (ROI) of 88.2%—still a sizable return. Since our experiment runs across non-advertising restaurants, the above calculation raises the possibility that some restaurants may have benefited from investing in advertising.

[23] According to accounting firm Baker Tilly (http://www.bakertilly.com/uploads/restaurant-benchmarking.pdf), the variable cost (food cost, i.e., the cost of goods sold) is roughly 28%–32% of total sales.

# Appendix E Heterogeneous Advertising Effects by Interactions of Business Attributes

We discuss the main takeaways from these results in Section 4.2. In this appendix, we report all heterogeneous effect tables based on interactions of business attributes and describe additional findings.

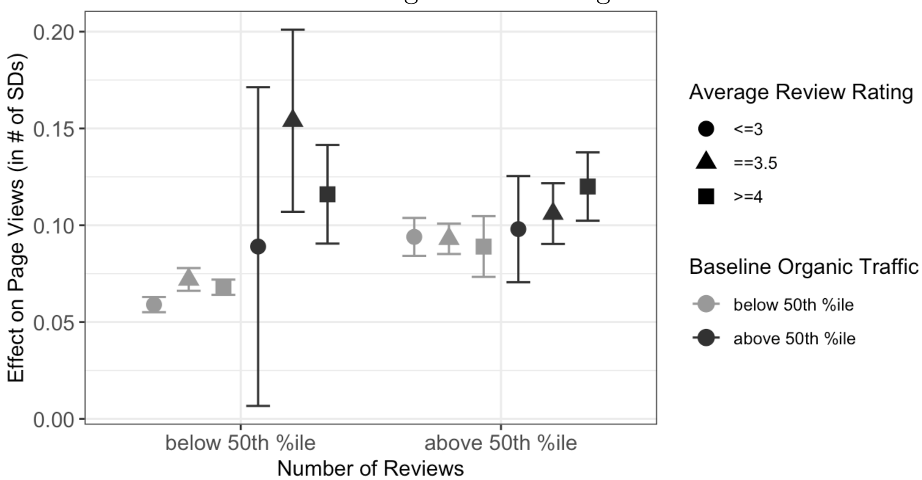
The figures and tables in this appendix are constructed in the following way:

- Figure E1 - E2 use different x-axis positions, point shapes, and colors to denote different restaurant attributes and show heterogeneous advertising effects based on different combinations of attributes (e.g., Figure E1 corresponds to Table E3 that shows heterogeneous effects based on ratings, number of reviews, and organic traffic).
- Table E1 - E9 report regression estimates of heterogeneous advertising effects based on different combinations of attributes.

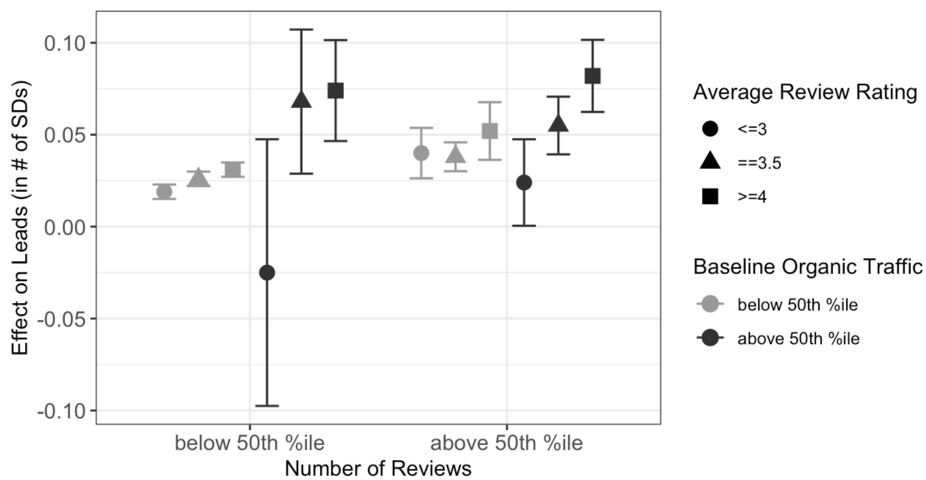
The heterogeneous effects based on the interactions of business attributes are useful in two ways—they show the combination of attributes that yield higher gains from advertising and check the robustness of the marginal contribution of each attribute on advertising gains conditional other attributes. Another way to look at these results is to examine whether a restaurant can make up for its shortcoming in one attribute by having an advantage in another attribute, which is related both to the relative importance of different attributes and to whether the attributes are complementary or substitutable in affecting advertising gains. In the example of rating and traffic (Appendix E Table E1), the advertising gains from higher ratings increase with traffic and vice versa, which leads to the observation that the gains for restaurants with the worst rating and highest traffic or the best rating and lowest traffic both result in lower advertising gains than restaurants with median ratings and median traffic. There are also substitutable attributes. For example, Appendix E Table E2 shows the advertising gains for restaurants with a different number of reviews and organic traffic. The marginal increase in advertising gains associated with more reviews declines with organic traffic and vice versa. The substitutability between number of reviews and organic traffic could be due to the main mechanism driving differential advertising effects through these two attributes being similar. For example, both are more or less measuring that restaurants that already have a popular presence on Yelp may gain more from advertising.

Figure E1: Heterogeneous Advertising Effects by Number of Reviews, Ratings, and Organic Traffic

A. Advertising Effects on Page Views



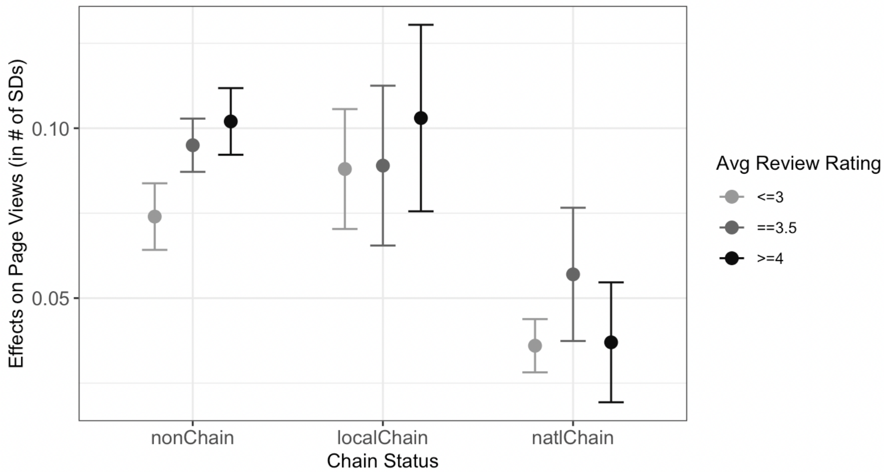

B. Advertising Effects on Leads



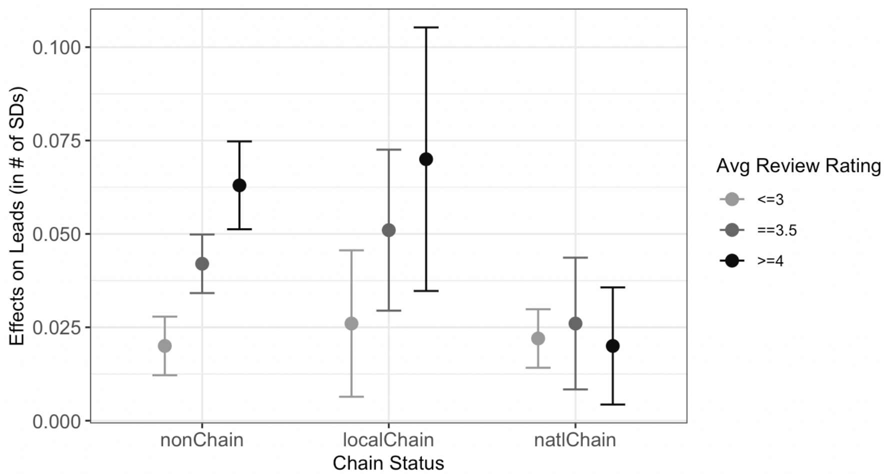

*Notes:* **1** The figures show heterogeneous advertising effects on standardized page views and leads for restaurants with different pre-experiment number of reviews (distinguished by the x-axis values), organic page views (distinguished by color), and mean review ratings (distinguished by shape). **2** The effects are obtained by running regression equation (2) in Appendix B and replacing variable Z by the exhaustive interactive terms of the dummy variables for discrete attributes. The regression result table is shown in Appendix E Table E3.

Figure E2: Heterogeneous Advertising Effects by Chain Status and Review Rating

A. Advertising Effects on Page Views




B. Advertising Effects on Leads




*Notes:* **1** The figures show heterogeneous advertising effects on standardized page views and leads for restaurants with different chain statuses (distinguished by x-axis values) and different ratings (distinguished by color). The error bars are the 95th confidence interval of the estimates. **2** The effects are obtained by running regression equation (2) in Appendix B and replacing variable Z by the cross-product terms of discrete attributes. The regression result table is shown in Appendix E Table E11.

Table E1: Heterogeneous Advertising Effects by Review Rating and Organic Traffic

*[Outcomes Standardized]*

| *Weighted* | (1) | (2) | (3) |
|---|---|---|---|
| *Regressors*[a] | Page Views | Leads | Reviews |
| ExprOn×Treated× | | | |
| (≤3 Stars)×(Traffic≤50th pctl) | 0.065*** | 0.022*** | 0.009 |
| | (0.002) | (0.002) | (0.007) |
| (≤3 Stars)×(Traffic∈(50,90]th pctl) | 0.088*** | 0.020* | 0.001 |
| | (0.009) | (0.009) | (0.026) |
| (≤3 Stars)×(Traffic>90th pctl) | 0.108 | -0.023 | 0.090 |
| | (0.108) | (0.085) | (0.166) |
| (3.5 Stars)×(Traffic≤50th pctl) | 0.076*** | 0.028*** | 0.015** |
| | (0.002) | (0.002) | (0.006) |
| (3.5 Stars)×(Traffic∈(50,90]th pctl) | 0.105*** | 0.053*** | 0.025* |
| | (0.006) | (0.006) | (0.013) |
| (3.5 Stars)×(Traffic>90th pctl) | 0.144*** | 0.077* | 0.063 |
| | (0.043) | (0.038) | (0.063) |
| (≥4 Stars)×(Traffic≤50th pctl) | 0.069*** | 0.033*** | 0.014* |
| | (0.002) | (0.002) | (0.006) |
| (≥4 Stars)×(Traffic∈(50,90]th pctl) | 0.112*** | 0.065*** | 0.045*** |
| | (0.005) | (0.006) | (0.010) |
| (≥4 Stars)×(Traffic>90th pctl) | 0.130*** | 0.117*** | 0.061 |
| | (0.025) | (0.030) | (0.044) |
| Other Controls and Fixed Effects | x | x | x |
| N | 218,266 | 218,266 | 218,266 |

* $p<0.05$, ** $p<0.01$, *** $p<0.001$. Standard errors are in parentheses, clustered at the business level.

[a] Regressors are weighted according to the regression equation (2) in Appendix B. We replaced variable $Z$ in equation (2) by the full set of mutually exclusive interaction terms of discrete restaurant attributes (their dummy variable representations). Coefficients for the interaction terms between ExprOn×Treated and the dummy variables can be interpreted as advertising effects for restaurants with the specified combination of restaurant attributes.

*Notes:* **1** The table shows heterogeneous advertising effects for businesses with different pre-experiment review ratings and organic traffic levels. **2** Due to Yelp's requirement for advertising businesses, most restaurants have at least a 3-star average rating. In the category of $I(\leq 3Stars)$, 95% of the restaurants have a 3-star average rating and 5% have a 2.5-star average rating. **3** Observations are at the business×month level. Business fixed effects and month fixed effects are added, and standard errors are clustered at the business level.

Table E2: Heterogeneous Advertising Effects by Review Number and Organic Traffic

*[Outcomes Standardized]*

| *Weighted* | (1) | (2) | (3) |
|---|---|---|---|
| *Regressors*[a] | Page Views | Leads | Reviews |
| ExprOn×Treated× | | | |
| (#Rev≤50th pctl)×(Traffic≤50th pctl) | 0.067*** | 0.026*** | 0.013*** |
| | (0.001) | (0.001) | (0.003) |
| (#Revv≤50th pctl)×(Traffic>50th pctl) | 0.116*** | 0.065*** | 0.052** |
| | (0.012) | (0.012) | (0.019) |
| (#Rev>50th pctl)×(Traffic≤50th pctl) | 0.092*** | 0.040*** | 0.015 |
| | (0.003) | (0.003) | (0.011) |
| (#Rev>50th pctl)×(Traffic>50th pctl) | 0.113*** | 0.067*** | 0.038*** |
| | (0.006) | (0.006) | (0.011) |
| Other Controls and Fixed Effects | x | x | x |
| N | 218,821 | 218,821 | 218,821 |

* $p<0.05$, ** $p<0.01$, *** $p<0.001$. Standard errors are in parentheses, clustered at the business level.

[a] Regressors are weighted according to the regression equation (2) in Appendix B. We replaced variable $Z$ in equation (2) by the full set of mutually exclusive interaction terms of discrete restaurant attributes (their dummy variable representations). Coefficients for the interaction terms between ExprOn×Treated and the dummy variables can be interpreted as advertising effects for restaurants with the specified combination of restaurant attributes.

*Notes:* **1** The table shows heterogeneous advertising effects for businesses with different pre-experiment review numbers and organic traffic. **2** Observations are at the business×month level. Business fixed effects and month fixed effects are added, and standard errors are clustered at the business level.

Table E3: Heterogeneous Advertising Effects by Review Rating, Review Number, and Organic Traffic

*[Outcomes Standardized]*

| *Weighted* *Regressors*[a] | (1) Page Views | (2) Leads | (3) Reviews |
|---|---|---|---|
| ExprOn×Treated | | | |
| (≤3 Stars)×(#Rev≤50th)×(Traffic≤50th) | 0.059*** | 0.019*** | 0.009 |
| | (0.002) | (0.002) | (0.007) |
| (≤3 Stars)×(#Rev≤50th)×(Traffic>50th) | 0.089* | -0.025 | 0.051 |
| | (0.042) | (0.037) | (0.077) |
| (≤3 Stars)×(#Rev>50th)×(Traffic≤50th) | 0.094*** | 0.040*** | 0.013 |
| | (0.005) | (0.007) | (0.025) |
| (≤3 Stars)×(#Rev>50th)×(Traffic>50th) | 0.098*** | 0.023 | 0.014 |
| | (0.014) | (0.012) | (0.030) |
| (=3.5 Stars)×(#Rev≤50th)×(Traffic≤50th) | 0.072*** | 0.026*** | 0.014** |
| | (0.003) | (0.002) | (0.006) |
| (=3.5 Stars)×(#Rev≤50th)×(Traffic>50th) | 0.154*** | 0.068*** | 0.046 |
| | (0.024) | (0.020) | (0.038) |
| (=3.5 Stars)×(#Rev>50th) ×(Traffic≤50th) | 0.093*** | 0.038*** | 0.018 |
| | (0.004) | (0.004) | (0.016) |
| (=3.5 Stars)×(#Rev>50th)×(Traffic>50th) | 0.106*** | 0.055*** | 0.029 |
| | (0.008) | (0.008) | (0.015) |
| (≥4 Stars)×(#Rev≤50th)×(Traffic≤50th) | 0.068*** | 0.031*** | 0.015** |
| | (0.002) | (0.002) | (0.006) |
| (≥4 Stars)×(#Rev≤50th)×(Traffic>50th) | 0.116*** | 0.074*** | 0.061** |
| | (0.013) | (0.014) | (0.023) |
| (≥4 Stars)×(#Rev>50th)×(Traffic≤50th) | 0.089*** | 0.052*** | 0.004 |
| | (0.008) | (0.008) | (0.023) |
| (≥4 Stars)×(#Rev>50th)×(Traffic>50th) | 0.120*** | 0.082*** | 0.048** |
| | (0.009) | (0.010) | (0.016) |
| Other Controls and Fixed Effects | x | x | x |
| N | 218,266 | 218,266 | 218,266 |

* $p<0.05$, ** $p<0.01$, *** $p<0.001$. Standard errors are in parentheses, clustered at the business level.

[a] Regressors are weighted according to the regression equation (2) in Appendix B. We replaced variable $Z$ in equation (2) by the full set of mutually exclusive interaction terms of discrete restaurant attributes (their dummy variable representations). Coefficients for the interaction terms between ExprOn×Treated and the dummy variables can be interpreted as advertising effects for restaurants with the specified combination of restaurant attributes.

*Notes:* **1** The table shows heterogeneous advertising effects for businesses with different pre-experiment review ratings, number of reviews, and organic traffic. The estimated heterogeneous effects are also plotted in Figure E1 for easier visualization. **2** Due to Yelp's requirement for advertising businesses, most restaurants have at least a 3-star average rating. In the category of $I(\leq 3Stars)$, 95% of the restaurants have a 3-star average rating and 5% have a 2.5-star average rating. **3** Observations are at the business×month level. Business fixed effects and month fixed effects are added, and standard errors are clustered at the business level.

Table E4: Heterogeneous Advertising Effects by Review Rating, Number of Reviews, and Chain Status

*[Outcomes Standardized]*

| *Weighted* *Regressors*[a] | (1) Page Views | (2) Leads | (3) Reviews |
|---|---|---|---|
| ExprOn×Treated× | | | |
| (≤3 Stars)×(#Rev≤50th)×non-natlChain | 0.063*** | 0.016*** | 0.011 |
| | (0.003) | (0.003) | (0.008) |
| (=3.5 Stars)×(#Rev≤50th)×non-natlChain | 0.084*** | 0.031*** | 0.018** |
| | (0.004) | (0.003) | (0.007) |
| (≥4 Stars)×(#Rev≤50th)×non-natlChain | 0.083*** | 0.043*** | 0.028*** |
| | (0.004) | (0.004) | (0.008) |
| (≤3 Stars)×(#Rev>50th)×non-natlChain | 0.096*** | 0.029*** | 0.014 |
| | (0.009) | (0.008) | (0.022) |
| (=3.5 Stars)×(#Rev>50th)×non-natlChain | 0.104*** | 0.052*** | 0.025* |
| | (0.007) | (0.006) | (0.012) |
| (≥4 Stars)×(#Rev>50th)×non-natlChain | 0.118*** | 0.080*** | 0.045** |
| | (0.008) | (0.010) | (0.015) |
| (≤3 Stars)×(#Rev≤50th)×natlChain | 0.032*** | 0.021*** | -0.005 |
| | (0.003) | (0.003) | (0.010) |
| (=3.5 Stars)×(#Rev≤50th)×natlChain | 0.032*** | 0.010** | -0.006 |
| | (0.003) | (0.003) | (0.013) |
| (≥4 Stars)×(#Rev≤50th)×natlChain | 0.026*** | 0.014*** | 0.009 |
| | (0.004) | (0.003) | (0.016) |
| (≤3 Stars)×(#Rev>50th)×natlChain | 0.087* | 0.024 | -0.081 |
| | (0.036) | (0.024) | (0.063) |
| (=3.5 Stars)×(#Rev>50th)×natlChain | 0.127*** | 0.070* | 0.136* |
| | (0.035) | (0.031) | (0.067) |
| (≥4 Stars)×(#Rev>50th)×natlChain | 0.090 | 0.036 | -0.054 |
| | (0.047) | (0.046) | (0.136) |
| Other Controls and FEs | x | x | x |
| N | 218,266 | 218,266 | 218,266 |

* $p<0.05$, ** $p<0.01$, *** $p<0.001$. Standard errors are in parentheses, clustered at the business level.

[a] Regressors are weighted according to the regression equation (2) in Appendix B. We replaced variable $Z$ in equation (2) by the full set of mutually exclusive interaction terms of discrete restaurant attributes (their dummy variable representations). Coefficients for the interaction terms between ExprOn×Treated and the dummy variables can be interpreted as advertising effects for restaurants with the specified combination of restaurant attributes.

*Notes:* **1** The table shows the heterogeneous advertising effects for businesses with different pre-experiment review ratings, number of reviews, and chain status. **2** Due to Yelp's requirement for advertising businesses, most restaurants have at least a 3-star average rating. In the category of $I(\leq 3Stars)$, 95% of the restaurants have a 3-star average rating and 5% have a 2.5-star average rating. **3** Observations are at the business×month level. Business fixed effects and month fixed effects are added, and standard errors are clustered at the business level.

Table E5: Heterogeneous Advertising Effects by Restaurant Age, Number of Reviews, and Chain Status

| *[Outcomes Standardized]* | | | |
|---|---|---|---|
| *Weighted* | (1) | (2) | (3) |
| *Regressors*[a] | Page Views | Leads | Reviews |
| ExprOn×Treated× | | | |
| (≤2Yrs)×(#Rev≤50th)×non-natlChain | 0.101*** | 0.059*** | 0.044** |
| | (0.010) | (0.010) | (0.017) |
| (>2Yrs)×(#Rev≤50th)×non-natlChain | 0.074*** | 0.028*** | 0.016*** |
| | (0.002) | (0.002) | (0.004) |
| (≤2Yrs)×(#Rev>50th)×non-natlChain | 0.130*** | 0.089* | 0.092 |
| | (0.031) | (0.035) | (0.049) |
| (>2Yrs)×(#Rev>50th)×non-natlChain | 0.109*** | 0.061*** | 0.029** |
| | (0.004) | (0.005) | (0.009) |
| (≤2Yrs)×(#Rev≤50th)×natlChain | 0.031** | 0.007 | -0.040 |
| | (0.010) | (0.010) | (0.030) |
| (>2Yrs)×(#Rev≤50th)×natlChain | 0.029*** | 0.015*** | -0.001 |
| | (0.002) | (0.002) | (0.007) |
| (≤2Yrs)×(#Rev>50th)×natlChain | -0.066 | -0.116* | -0.093 |
| | (0.064) | (0.057) | (0.128) |
| (>2Yrs)×(#Rev>50th)×natlChain | 0.101*** | 0.057** | 0.072 |
| | (0.015) | (0.019) | (0.048) |
| Other Controls and FEs | x | x | x |
| N | 218,821 | 218,821 | 218,821 |

* $p<0.05$, ** $p<0.01$, *** $p<0.001$. Standard errors are in parentheses, clustered at the business level.

[a] Regressors are weighted according to the regression equation (2) in Appendix B. We replaced variable $Z$ in equation (2) by the full set of mutually exclusive interaction terms of discrete restaurant attributes (their dummy variable representations). The coefficients for the interaction terms between ExprOn×Treated and the dummy variables can be interpreted as advertising effects for restaurants with the specified combination of restaurant attributes.

*Notes:* **1** The table shows heterogeneous advertising effects for businesses with different pre-experiment age, number of reviews, and chain status. **2** Observations are at the business×month level. Business fixed effects and zip code fixed effects are added, and standard errors are clustered at the business level.

Table E6: Heterogeneous Advertising Effects by Review Rating, Age, and Chain Status

*[Outcomes Standardized]*

| *Weighted* | (1) | (2) | (3) |
|---|---|---|---|
| *Regressors*[a] | Page Views | Leads | Reviews |
| ExprOn×Treated× | | | |
| ($\leq$3 Stars)×(Age$\leq$2Yrs)×non-natlChain | 0.080** | 0.004 | 0.069 |
| | (0.026) | (0.017) | (0.052) |
| ($\leq$3 Stars)×(Age$>$2Yrs)×non-natlChain | 0.077*** | 0.023*** | 0.010 |
| | (0.003) | (0.003) | (0.009) |
| (=3.5 Stars)×(Age$\leq$2Yrs)×non-natlChain | 0.125*** | 0.062*** | 0.052 |
| | (0.018) | (0.017) | (0.032) |
| (=3.5 Stars)×(Age$>$2Yrs)×non-natlChain | 0.091*** | 0.040*** | 0.019* |
| | (0.004) | (0.004) | (0.007) |
| ($\geq 4Stars$)×(Age$\leq$2Yrs)×non-natlChain | 0.109*** | 0.080*** | 0.060* |
| | (0.017) | (0.019) | (0.026) |
| ($\geq$4 Stars)×(Age$>$2Yrs)×non-natlChain | 0.100*** | 0.059*** | 0.031*** |
| | (0.004) | (0.005) | (0.009) |
| ($\leq$3 Stars)×(Age$\leq$2Yrs)×natlChain | 0.081*** | 0.028 | -0.177*** |
| | (0.024) | (0.025) | (0.030) |
| ($\leq$3 Stars)×(Age$>$2Yrs)×natlChain | 0.030*** | 0.018*** | -0.005 |
| | (0.003) | (0.003) | (0.010) |
| (=3.5 Stars)×(Age$\leq$2Yrs)×natlChain | 0.179 | 0.105 | 0.234 |
| | (0.161) | (0.102) | (0.207) |
| (=3.5 Stars)×(Age$>$2Yrs)×natlChain | 0.051*** | 0.023** | 0.022 |
| | (0.006) | (0.008) | (0.020) |
| ($\geq$4 Stars)×(Age$\leq$2Yrs)×natlChain | 0.028 | -0.002 | 0.016 |
| | (0.027) | (0.016) | (0.077) |
| ($\geq$4 Stars)×(Age$>$2Yrs)×natlChain | 0.041*** | 0.027*** | 0.010 |
| | (0.006) | (0.007) | (0.025) |
| Other Controls and FEs | x | x | x |
| N | 218,230 | 218,230 | 218,230 |

* $p<0.05$, ** $p<0.01$, *** $p<0.001$. Standard errors are in parentheses, clustered at the business level.

[a] Regressors are weighted according to the regression equation (2) in Appendix B. We replaced variable $Z$ in equation (2) by the full set of mutually exclusive interaction terms of discrete restaurant attributes (their dummy variable representations). Coefficients for the interaction terms between ExprOn×Treated and the dummy variables can be interpreted as advertising effects for restaurants with the specified combination of restaurant attributes.

*Notes:* **1** The table shows heterogeneous advertising effects for businesses with different pre-experiment mean review ratings, restaurant age, and chain status. **2** Due to Yelp's requirement for advertising businesses, most restaurants have at least a 3-star average rating. In the category of $I(\leq 3Stars)$, 95% of the restaurants have a 3-star average rating and 5% have a 2.5-star average rating. **3** Observations are at the business×month level. Business fixed effects and month fixed effects are added, and standard errors are clustered at the business level.

Table E7: Heterogeneous Advertising Effects by Organic Traffic and Chain Status

*[Outcomes Standardized]*

| *Weighted* <br> *Regressors*[a] | (1) <br> Page Views | (2) <br> Leads | (3) <br> Reviews |
|---|---|---|---|
| ExprOn×Treated | | | |
| (Traffic≤25th pctl)×non-natlChain | 0.060*** | 0.024*** | 0.019*** |
| | (0.002) | (0.001) | (0.004) |
| (Traffic∈(25,50]th pctl)×non-natlChain | 0.088*** | 0.035*** | 0.010 |
| | (0.002) | (0.002) | (0.006) |
| (Traffic∈(50,75]th pctl)×non-natlChain | 0.106*** | 0.051*** | 0.028** |
| | (0.004) | (0.004) | (0.008) |
| (Traffic∈(75,90]th pctl)×non-natlChain | 0.107*** | 0.064*** | 0.041** |
| | (0.007) | (0.008) | (0.015) |
| (Traffic∈>90th pctl)×non-natlChain | 0.132*** | 0.102*** | 0.061 |
| | (0.021) | (0.024) | (0.037) |
| (Traffic≤25th pctl)×natlChain | 0.027*** | 0.015*** | -0.002 |
| | (0.002) | (0.002) | (0.007) |
| (Traffic∈(25,50]th pctl)×natlChain | 0.066*** | 0.019* | 0.042 |
| | (0.007) | (0.009) | (0.042) |
| (Traffic∈(50,75]th pctl)×natlChain | 0.093*** | 0.040* | 0.017 |
| | (0.014) | (0.018) | (0.059) |
| (Traffic∈(75,90]th pctl)×natlChain | 0.230*** | 0.240** | 0.193 |
| | (0.053) | (0.086) | (0.141) |
| (Traffic∈>90th pctl)×natlChain | 0.006 | -0.086 | 0.029 |
| | (0.139) | (0.146) | (0.319) |
| Other Controls and FEs | x | x | x |
| N | 218,821 | 218,821 | 218,821 |

* p<0.05, ** p<0.01, *** p<0.001. Standard errors are in parentheses, clustered at the business level.

[a] Regressors are weighted according to the regression equation (2) in Appendix B. We replaced variable $Z$ in equation (2) by the full set of mutually exclusive interaction terms of discrete restaurant attributes (their dummy variable representations). Coefficients for the interaction terms between ExprOn×Treated and the dummy variables can be interpreted as advertising effects for restaurants with the specified combination of restaurant attributes.

*Notes:* **1** The table shows heterogeneous advertising effects for businesses with different pre-experiment organic traffic and chain status. **2** Observations are at the business×month level. Business fixed effects and month fixed effects are added, and standard errors are clustered at the business level.

Table E8: Heterogeneous Advertising Effects by Review Rating and Price Level

*[Outcomes Standardized]*

| *Weighted* | (1) | (2) | (3) |
|---|---|---|---|
| *Regressors*[a] | Page Views | Leads | Reviews |
| ExprOn×Treated× | | | |
| (≤3 Stars)×$ | 0.066*** | 0.025*** | 0.003 |
| | (0.003) | (0.003) | (0.010) |
| (≤3 Stars)×$$ | 0.071*** | 0.018*** | 0.014 |
| | (0.005) | (0.004) | (0.013) |
| (≤3 Stars)×$$$ | 0.096* | 0.010 | 0.018 |
| | (0.041) | (0.032) | (0.055) |
| (≤3 Stars)×$$$$ | 0.113** | 0.055* | -0.102 |
| | (0.043) | (0.023) | (0.086) |
| (=3.5 Stars)×$ | 0.075*** | 0.029*** | 0.028** |
| | (0.004) | (0.004) | (0.009) |
| (=3.5 Stars)×$$ | 0.095*** | 0.045*** | 0.022* |
| | (0.005) | (0.005) | (0.010) |
| (=3.5 Stars)×$$$ | 0.115*** | 0.054*** | 0.007 |
| | (0.018) | (0.013) | (0.024) |
| (=3.5 Stars)×$$$$ | 0.080 | 0.001 | 0.060 |
| | (0.045) | (0.030) | (0.098) |
| (≥4 Stars)×$ | 0.085*** | 0.050*** | 0.029** |
| | (0.005) | (0.006) | (0.010) |
| (≥4 Stars)×$$ | 0.103*** | 0.068*** | 0.046*** |
| | (0.007) | (0.008) | (0.014) |
| (≥4 Stars)×$$$ | 0.121*** | 0.071*** | 0.033 |
| | (0.015) | (0.017) | (0.023) |
| (≥4 Stars)×$$$$ | 0.086* | 0.017 | -0.034 |
| | (0.036) | (0.032) | (0.057) |
| Other Controls and FEs | x | x | x |
| N | 216,166 | 216,166 | 216,166 |

* $p<0.05$, ** $p<0.01$, *** $p<0.001$. Standard errors are in parentheses, clustered at the business level.

[a] Regressors are weighted according to the regression equation (2) in Appendix B. We replaced variable $Z$ in equation (2) by the full set of mutually exclusive interaction terms of discrete restaurant attributes (their dummy variable representations). Coefficients for the interaction terms between ExprOn×Treated and the dummy variables can be interpreted as advertising effects for restaurants with the specified combination of restaurant attributes.

*Notes:* **1** The table shows heterogeneous advertising effects for businesses with different pre-experiment review ratings and price levels. **2** Due to Yelp's requirement for advertising businesses, most restaurants have at least a 3-star average rating. In the category of $I(\leq 3Stars)$, 95% of the restaurants have a 3-star average rating and 5% have a 2.5-star average rating. **3** Observations are at the business×month level. Restaurants without price information are dropped from the regression. Business fixed effects and zip code fixed effects are added, and standard errors are clustered at the business level.

Table E9: Heterogeneous Advertising Effects by Organic Traffic and Price Level

*[Outcomes Standardized]*

| *Weighted* | (1) | (2) | (3) |
|---|---|---|---|
| *Regressors*[a] | Page Views | Leads | Reviews |
| ExprOn×Treated× | | | |
| (Traffic≤50th pctl)×$ | 0.065*** | 0.028*** | 0.015*** |
| | (0.002) | (0.002) | (0.004) |
| (Traffic≤50th pctl)×$$ | 0.071*** | 0.027*** | 0.011 |
| | (0.002) | (0.002) | (0.006) |
| (Traffic≤50th pctl)×$$$ | 0.112*** | 0.040*** | 0.018 |
| | (0.006) | (0.005) | (0.015) |
| (Traffic≤50th pctl)×$$$$ | 0.114*** | 0.052*** | 0.087* |
| | (0.015) | (0.011) | (0.037) |
| (Traffic∈(50,90]th pctl)×$ | 0.100*** | 0.058*** | 0.040* |
| | (0.008) | (0.009) | (0.017) |
| (Traffic∈(50,90]th pctl)×$$ | 0.103*** | 0.056*** | 0.033** |
| | (0.005) | (0.005) | (0.010) |
| (Traffic∈(50,90]th pctl)×$$$ | 0.124*** | 0.057*** | 0.035* |
| | (0.008) | (0.008) | (0.016) |
| (Traffic∈(50,90]th pctl)×$$$$ | 0.130*** | 0.053** | -0.002 |
| | (0.020) | (0.017) | (0.040) |
| (Traffic>90th pctl)×$ | 0.212*** | 0.101 | 0.109 |
| | (0.063) | (0.065) | (0.107) |
| (Traffic>90th pctl)×$$ | 0.136*** | 0.108*** | 0.090 |
| | (0.026) | (0.031) | (0.050) |
| (Traffic>90th pctl)×$$$ | 0.121** | 0.107* | 0.009 |
| | (0.046) | (0.049) | (0.063) |
| (Traffic>90th pctl)×$$$$ | 0.001 | -0.071 | -0.077 |
| | (0.087) | (0.084) | (0.153) |
| Other Controls and FEs | x | x | x |
| N | 216,592 | 216,592 | 216,592 |

* $p<0.05$, ** $p<0.01$, *** $p<0.001$. Standard errors are in parentheses, clustered at the business level.

[a] Regressors are weighted according to the regression equation (2) in Appendix B. We replaced variable $Z$ in equation (2) by the full set of mutually exclusive interaction terms of discrete restaurant attributes (their dummy variable representations). Coefficients for the interaction terms between ExprOn×Treated and the dummy variables can be interpreted as advertising effects for restaurants with the specified combination of restaurant attributes.

*Notes:* **1** The table shows heterogeneous advertising effects for businesses with different pre-experiment organic page views and price levels. **2** Observations are at the business×month level. Restaurants without price information are dropped from the regression. Business fixed effects and zip code fixed effects are added, and standard errors are clustered at the business level.

Table E10: Heterogeneous Advertising Effects by Review Rating, Organic Traffic, and Price Level

*[Outcomes Standardized]*

| *Weighted* | (1) | (2) | (3) |
|---|---|---|---|
| *Regressors*[a] | Page Views | Leads | Reviews |
| ExprOn×Treated× | | | |
| (≤3 Stars)×(Traffic≤50th)×($,$$) | 0.064*** | 0.022*** | 0.009 |
| | (0.002) | (0.002) | (0.007) |
| (≤3 Stars)×(Traffic≤50th)×($$$,$$$$) | 0.089*** | 0.031*** | 0.018 |
| | (0.011) | (0.008) | (0.041) |
| (≤3 Stars)×(Traffic>50th)×($,$$) | 0.090*** | 0.019 | 0.010 |
| | (0.011) | (0.011) | (0.030) |
| (≤3 Stars)×(Traffic>50th)×($$$,$$$$) | 0.114* | 0.012 | 0.010 |
| | (0.056) | (0.045) | (0.077) |
| (=3.5 Stars)×(Traffic≤50th)×($,$$) | 0.073*** | 0.028*** | 0.015* |
| | (0.002) | (0.002) | (0.006) |
| (=3.5 Stars)×(Traffic≤50th)×($$$,$$$$) | 0.118*** | 0.034*** | 0.014 |
| | (0.008) | (0.006) | (0.022) |
| (=3.5 Stars)×(Traffic>50th)×($,$$) | 0.108*** | 0.055*** | 0.035* |
| | (0.008) | (0.008) | (0.016) |
| (=3.5 Stars)×(Traffic>50th)×($$$,$$$$) | 0.116*** | 0.061*** | 0.017 |
| | (0.023) | (0.018) | (0.032) |
| (≥4 Stars)×(Traffic≤50th)×($,$$) | 0.066*** | 0.032*** | 0.013* |
| | (0.002) | (0.002) | (0.006) |
| (≥4 Stars)×(Traffic≤50th)×($$$,$$$$) | 0.111*** | 0.048*** | 0.024 |
| | (0.009) | (0.007) | (0.019) |
| (≥4 Stars)×(Traffic>50th)×($,$$) | 0.118*** | 0.083*** | 0.060*** |
| | (0.008) | (0.010) | (0.016) |
| (≥4 Stars)×(Traffic>50th)×($$$,$$$$) | 0.121*** | 0.069*** | 0.026 |
| | (0.016) | (0.018) | (0.025) |
| Other Controls and FEs | x | x | x |
| N | 216,202 | 216,202 | 216,202 |

* p<0.05, ** p<0.01, *** p<0.001. Standard errors are in parentheses, clustered at the business level.

[a] Regressors are weighted according to the regression equation (2) in Appendix B. We replaced variable $Z$ in equation (2) by the full set of mutually exclusive interaction terms of discrete restaurant attributes (their dummy variable representations). Coefficients for the interaction terms between ExprOn×Treated and the dummy variables can be interpreted as advertising effects for restaurants with the specified combination of restaurant attributes.

*Notes:* **1** The table shows heterogeneous advertising effects for businesses with different pre-experiment ratings, organic page views, and price levels. **2** Due to Yelp's requirement for advertising businesses, most restaurants have at least a 3-star average rating. In the category of $I(\leq 3Stars)$, 95% of the restaurants have a 3-star average rating and 5% have a 2.5-star average rating. **3** Observations are at the business×month level. Restaurants without price information are dropped from the regression. Business fixed effects and month fixed effects are added, and standard errors are clustered at the business level.

Table E11: Heterogeneous Advertising Effects by Review Rating and Chain Status

*[Outcomes Standardized]*

| *Weighted* | (1) | (2) | (3) |
|---|---|---|---|
| *Regressors*[a] | Page Views | Leads | Reviews |
| ExprOn×Treated | | | |
| (≤3 Stars)×Independent | 0.074*** | 0.020*** | 0.001 |
| | (0.005) | (0.004) | (0.010) |
| (≤3 Stars)×localChain | 0.088*** | 0.026** | 0.085** |
| | (0.009) | (0.010) | (0.029) |
| (≤3 Stars)×natlChain | 0.036*** | 0.022*** | -0.008 |
| | (0.004) | (0.004) | (0.010) |
| (=3.5 Stars)×Independent | 0.095*** | 0.042*** | 0.023** |
| | (0.004) | (0.004) | (0.008) |
| (=3.5 Stars)×localChain | 0.089*** | 0.051*** | 0.003 |
| | (0.012) | (0.011) | (0.029) |
| (=3.5 Stars)×natlChain | 0.057*** | 0.026** | 0.032 |
| | (0.010) | (0.009) | (0.021) |
| (≥4 Stars)×Independent | 0.102*** | 0.063*** | 0.037*** |
| | (0.005) | (0.006) | (0.009) |
| (≥4 Stars)×localChain | 0.103*** | 0.070*** | 0.037 |
| | (0.014) | (0.018) | (0.030) |
| (≥4 Stars)×natlChain | 0.037*** | 0.020* | 0.004 |
| | (0.009) | (0.008) | (0.026) |
| Other Controls and FEs | x | x | x |
| N | 218,266 | 218,266 | 218,266 |

* $p<0.05$, ** $p<0.01$, *** $p<0.001$. Standard errors are in parentheses, clustered at the business level.

[a] Regressors are weighted according to the regression equation (2) in Appendix B. We replaced variable $Z$ in equation (2) by the full set of mutually exclusive interaction terms of discrete restaurant attributes (their dummy variable representations). The coefficients for the interaction terms between ExprOn×Treated and the dummy variables can be interpreted as the advertising effects for the restaurants with a specific combination of restaurant attributes.

*Notes:* **1** The table shows the heterogeneous advertising effects for businesses with different pre-experiment average review ratings and chain statuses. The estimates are plotted in Appendix E Figure E2. **2** Due to Yelp's requirement for advertising businesses, most restaurants have at least a 3-star average rating. In the category of $I(\leq 3Stars)$, 95% of the restaurants have a 3-star average rating and 5% have a 2.5-star average rating. **2** The observations are at the business×month level. Business fixed effects and month fixed effects are added, and the standard errors are clustered at the business level.

# Appendix F Heterogeneous Advertising Effects on Conversion Rates

In the main paper, we present results on the heterogeneous effects of advertising directly on purchase intention (demand) measures: total page views, leads, number of reviews, and number of take-out orders.

In this section, we complement the results shown in the main paper by showing the effects of advertising on CTR and the conversion rate from page views to leads (CTA conversion rate). These results provide insights into the mechanism of how restaurant attributes affect the conversion of customers at different stages of the conversion funnel. Also, since the effects on final outcomes depend on paid and organic impressions, which differ widely among different restaurants, we condition on paid impressions (when examining CTR) and condition on page visits (when examining the CTA conversion rate) so that the results are less noisy.

Overall, the results are consistent with the findings when using purchase intentions as outcomes: effects are greater for independent restaurants, newer restaurants, and restaurants with better review ratings.

Table F1: Restaurant Attributes and Differences in Paid CTR

| *Outcome*: Standardized Paid CTR | ***Panel A.*** *Full Sample of Treated Restaurants* | ***Panel B.*** *Subsample of Treated Restaurants with Different Ratings* | | |
|---|---|---|---|---|
| | | (1) (<3.5 stars) | (2) (3.5 stars) | (3) (≥4 stars) |
| $Age(Year)\|Age<10$ | -0.006* | -0.032*** | -0.006 | -0.012** |
| | (0.003) | (0.006) | (0.004) | (0.005) |
| $I(Age \geq 10Yrs)$ | -0.058** | -0.23*** | -0.083** | -0.102*** |
| | (0.02) | (0.045) | (0.03) | (0.033) |
| $I(LocalChain)$ | -0.135*** | -0.101* | -0.013 | -0.322*** |
| | (0.027) | (0.046) | (0.037) | (0.052) |
| $I(NationalChain)$ | -0.421*** | -0.472*** | -0.258*** | -0.769*** |
| | (0.037) | (0.054) | (0.051) | (0.086) |
| $I(3.5star)$ | 0.094*** | | | |
| | (0.02) | | | |
| $I(4star)$ | 0.452*** | | | |
| | (0.01992) | | | |
| $I(\geq 4.5star)$ | 0.956*** | | | |
| | (0.027) | | | |
| $Ln(\#Reviews)$ | 0.034*** | -0.045** | 0.004 | 0.057*** |
| | (0.007) | (0.016) | (0.011) | (0.011) |
| Price category, subsample dummies | × | × | × | × |
| N | 20,835 | 3,087 | 7,624 | 10,124 |

Standard errors are in parentheses. * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

*Note:* **1** Panel A reports the results of regressing the treated restaurants' CTR of paid links on restaurant attributes. The sample includes all treated restaurants during the experiment period. Restaurant attributes are measured at the start of the experiment. Each observation is a business×month, and the paid CTR at the business×month level is calculated using the total number of paid page views divided by the total number of paid impressions (the number of times the business is shown in the paid search link) during each of the experiment month. The outcome variable paid CTR is standardized by its standard deviation in the regression sample. **2** We run the same regressions as in Panel A separately for restaurants with different review ratings to examine whether having more reviews always leads to higher paid CTRs for the restaurants. The results are shown in Panel B. **3** The excluded dummy variables are for restaurants that have been open for fewer than 10 years, independent restaurants, and restaurants with review ratings of three stars or below.

Table F2: Heterogeneous Click-through Rate of Paid Links, Alternative Controls

| *[Outcome Standardized]* | Paid CTR (Overall) |
|---|---|
| $Age(Year)\|Age < 10$ | -0.002 |
| | (0.003) |
| $I(Age \geq 10Yrs)$ | -0.027 |
| | (0.020) |
| $I(LocalChain)$ | -0.122*** |
| | (0.027) |
| $I(NationalChain)$ | -0.442*** |
| | (0.036) |
| $I(3.5star)$ | 0.101*** |
| | (0.019) |
| $I(4star)$ | 0.435*** |
| | (0.02017) |
| $I(\geq 4.5star)$ | 0.920*** |
| | (0.027) |
| $TrafficTier2$ | -0.030 |
| | (0.017) |
| $TrafficTier3$ | 0.004 |
| | (0.02) |
| $TrafficTier4$ | 0.192*** |
| | (0.024) |
| $TrafficTier5$ | 0.901*** |
| | (0.059) |
| $I(\$\$)$ | 0.200*** |
| | (0.017) |
| $I(\$\$\$)$ | 0.061* |
| | (0.026) |
| $I(\$\$\$\$)$ | -0.048 |
| | (0.049) |
| $I(\$missing)$ | -0.031 |
| | (0.084) |
| Subsample dummies | x |
| N | 20,835 |

Standard errors are in parentheses. * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

*Note:* **1** This table reports the results from OLS regressions of an advertising business's click-through rate (CTR) of paid links on predetermined restaurant attributes before the experiment. The sample includes all treated restaurants during the experiment period. The observation is at the business month level. **2** The dependent variable is the CTR of paid links of the treated businesses standardized by the standard deviation of the outcomes of the treated businesses during the treated period. **3** The dummy control variables exclude the lowest within each attributes category. We define tier 2 to tier 5 websites in terms of website traffic if the website's total page views before the experiment is in the 0-50th, 50-75th, 75-90th, 90–99th, or above 99th percentile.

Table F3: Heterogeneous Treatment Effects on Call-to-Action Conversion Rates by Restaurant Age and Chain Status

***Panel A.*** *Effect of Advertising on Overall CTA Conversion Rates*

*[Standardized Outcomes]*

| *Weighted Regressors*[a] | Conversion Rate | *Weighted Regressors*[a] | Conversion Rate |
|---|---|---|---|
| ExprOn×Treated | | ExprOn×Treated | |
| ×I(age<1yr) | -0.206*** | ×I(non-chain) | -0.265*** |
| | (0.04) | | (0.008) |
| ×I(age∈[1,2]yr) | -0.216*** | ×I(local chain) | -0.468*** |
| | (0.027) | | (0.04) |
| ×I(age∈[3,4]yr) | -0.279*** | ×I(national chain) | -0.854*** |
| | (0.019) | | (0.093) |
| ×I(age∈[5,10]yr) | -0.293*** | | |
| | (0.013) | | |
| ×I(age>10yr) | -0.385*** | | |
| | (0.021) | | |
| Other controls | x | | x |
| Fixed effects | Business, month | | Business, month |
| N | 218,670 | | 218,670 |
| $F$-test | $p$-value | | $p$-value |
| *Coeff1=Coeff2* | 0.81 | *Coeff1=Coeff2* | $4.59\times10^{-7}$ |
| *Coeff2=Coeff3* | 0.06 | *Coeff2=Coeff3* | $1.32\times10^{-4}$ |
| *Coeff3=Coeff4* | 0.56 | | |
| *Coeff4=Coeff5* | $2.54\times10^{-4}$ | | |

* $p<0.05$, ** $p<0.01$, *** $p<0.001$. Standard errors are in parentheses, clustered at the business level.

[a] Regressors are weighted according to regression equation (2) in Appendix B. The coefficient of the cross product of ExprOn×Treated and the dummy variable for each categorial level of the attribute represents the heterogeneous treatment effects. We have full interaction terms with all levels of each dimension of an attribute, so the coefficients can be directly interpreted as the advertising effect of each type of business. The coefficients can be directly compared to each other to show what type of businesses experience a bigger impact on the overall call-to-action conversion rate.

***Panel B.*** *Changes in CTA Conversion Rates (Paid vs. Organic Page Views)*

| $(\rho_{paid} - \rho_{organic})/\rho_{organic}$ | | | |
|---|---|---|---|
| Opens for | | Chain | |
| <1 year | -35.5% | non-chain | -43.5% |
| 1-2 years | -35.7% | local chain | -55.3% |
| 3-4 years | -45.6% | national chain | -53.4% |
| 5-10 years | -45.5% | | |
| >10 years | -49.8% | | |

*Note:* **1** Panel A reports the regression results of equation (2) in Appendix B that examine the advertising effect on the overall CTA conversion rate for restaurants with different attributes. The first column of panel A reports the regression result for restaurants across age bands, and the second column reports the regression result for restaurants with different chain statuses. **2** Panel B reports the change in the CTA conversion rate during the advertising period, comparing page visits from paid clicks and those from organic clicks. Letting $\rho_{org}$ and $\rho_{paid}$ denote the CTA conversion rate of organic and paid page views, respectively, the percentage change in the CTA conversion rate (paid vs. organic) measures $(\rho_{paid} - \rho_{org})/\rho_{org}$. **3** The regression is run at the business×month level including the full sample of businesses. It also estimates the heterogeneous effects on the overall CTA conversion rate in the two months after the experiment, and all the coefficients are small and statistically insignificant.

Table F4: Heterogeneous Treatment Effects on Call-to-Action Conversion Rates by Ratings and the Number of Reviews

***Panel A.*** *Effect of Advertising on Overall CTA Conversion Rates*

*[Outcomes Standardized]*

| *Weighted Regressors*[a] | Conversion Rate | *Weighted Regressors*[a] | Conversion Rate |
|---|---|---|---|
| ExprOn×Treated | | ExprOn×Treated | |
| ×I(<3.5 stars) | -0.488*** | ×I(#Rev in tier 1) | -0.527*** |
| | (0.031) | | (0.031) |
| ×I(3.5 stars) | -0.346*** | ×I(#Rev in tier 2) | -0.361*** |
| | (0.015) | | (0.015) |
| ×I(4 stars) | -0.214*** | ×I(#Rev in tier 3) | -0.228*** |
| | (0.012) | | (0.012) |
| ×I(>4 stars) | -0.224*** | ×I(#Rev in tier 4) | -0.116*** |
| | (0.029) | | (0.009) |
| Other controls | x | | x |
| Fixed effects | Business, month | | Business, month |
| N | 218,239 | | 218,670 |
| F-test | *p*-value | | p-value |
| *Coeff1=Coeff2* | $3.33\times10^{-5}$ | | $1.65\times10^{-6}$ |
| *Coeff2=Coeff3* | $1.92\times10^{-11}$ | | $1.01\times10^{-11}$ |
| *Coeff3=Coeff4* | 0.75 | | $1.07\times10^{-13}$ |

* $p<0.05$, ** $p<0.01$, *** $p<0.001$. Standard errors are in parentheses, clustered at the business level.

[a] Regressors are weighted according to regression equation (2) in Appendix B. The coefficient of the cross product of ExprOn×Treated and the dummy variable for each categorial level of the attribute represents the heterogeneous treatment effects. We have full interaction terms with all levels of each dimension of an attribute, so the coefficients can be directly interpreted as the advertising effect of each type of business. The coefficients can be directly compared to each other to show what type of businesses experience a bigger impact on the overall call-to-action (CTA) conversion rate.

***Panel B.*** *Changes in CTA Conversion Rates (Paid vs. Organic Page Views)*

| $(\rho_{paid} - \rho_{organic})/\rho_{organic}$ | | | |
|---|---|---|---|
| Rating | | # of Reviews | |
| <3.5 stars | -61.4% | Tier 1 (in 0-25th percentile) | -48.8% |
| 3.5 stars | -51.1% | Tier 2 (in 25-50th percentile) | -50.3% |
| 4 stars | -37.7% | Tier 3 (in 50-75th percentile) | -42.3% |
| >4 stars | -32.7% | Tier 4 (in 75-100th percentile) | -36.6% |

*Note:* **1** We conduct the same analysis as in Table F3 to produce this table, only changing the restaurant attributes examined. Panel A reports the regression results of equation (2) that examine the advertising effect on the overall CTA conversion rate for restaurants with different attributes. The first column of panel A reports the regression result for restaurants with different ratings, and the second column reports the regression result for restaurants with different volumes of reviews. We rank restaurant review volumes and create tiers based on the percentiles. **2** Panel B reports the percentage change in the CTA conversion rate during the advertising period, comparing page visits from paid clicks and those from organic clicks. Letting $\rho_{org}$ and $\rho_{paid}$ denote the CTA conversion rate of organic and paid page views, respectively, the percentage change in the CTA conversion rate (paid vs. organic) measures $(\rho_{paid} - \rho_{org})/\rho_{org}$. **3** The regression is run at the business×month level including the full sample of businesses. The differences in the number of observations are due to the few businesses for which we cannot match the review rating information. The regression also estimates the heterogeneous effects on the overall CTA conversion rate in the two months after the experiment, and all the coefficients are small and statistically insignificant.

Table F5: Heterogeneous Treatment Effects on Call-to-Action Conversion Rates by Price and Organic Traffic Ranking

***Panel A.*** *Effect of Paid Search Ads on Overall CTA Conversion Rates*

*[Outcomes Standardized]*

| *Weighted Regressors*[a] | Conversion Rate | *Weighted Regressors*[a] | Conversion Rate |
|---|---|---|---|
| ExprOn×Treated | | ExprOn×Treated | |
| ×I($) | -0.413*** | ×I(Traffic in tier 1) | -0.468*** |
| | (0.022) | | (0.017) |
| ×I($$) | -0.273*** | ×I(Traffic in tier 2) | -0.22*** |
| | (0.011) | | (0.011) |
| ×I($$$) | -0.192*** | ×I(Traffic in tier 3) | -0.087*** |
| | (0.013) | | (0.013) |
| ×I($$$$) | -0.159*** | ×I(Traffic in tier 4) | -0.055*** |
| | (0.035) | | (0.012) |
| ×I($missing) | -0.465*** | ×I(Traffic in tier 5) | 0.035 |
| | (0.156) | | (0.025) |
| Other controls | x | | x |
| Fixed effects | Business, month | | Business, month |
| N | 218,670 | | 218,670 |
| F-test | $p$-value | | $p$-value |
| *Coeff1=Coeff2* | $1.25\times10^{-8}$ | *Coeff1=Coeff2* | $2.83\times10^{-33}$ |
| *Coeff2=Coeff3* | $3.18\times10^{-6}$ | *Coeff2=Coeff3* | $1.58\times10^{-14}$ |
| *Coeff3=Coeff4* | 0.378 | *Coeff3=Coeff4* | 0.072 |
| *Coeff4=Coeff5* | 0.056 | *Coeff4=Coeff5* | $1.41\times10^{-3}$ |

* $p<0.05$, ** $p<0.01$, *** $p<0.001$. Standard errors are in parentheses, clustered at the business level.

[a] Regressors are weighted according to regression equation (2). The coefficient of the cross product of ExprOn×Treated and the dummy variable for each categorial level of the attribute represents the heterogeneous treatment effects. We have full interaction terms with all levels of each dimension of an attribute, so the coefficients can be directly interpreted as the advertising effect of each type of business. The coefficients can be directly compared to each other to show what type of businesses experience a bigger impact on the overall call-to-action (CTA) conversion rate.

***Panel B.*** *Changes in CTA Conversion Rates (Paid vs. Organic Page Views)*

| $(\rho_{paid} - \rho_{organic})/\rho_{organic}$ | | | |
|---|---|---|---|
| Price | | Pre-Experiment Organic Traffic | |
| $ | -47.8% | Tier 1 (in 0-50th percentile) | -50.2% |
| $$ | -45.0% | Tier 2 (in 50-75th percentile) | -42.3% |
| $$$ | -44.4% | Tier 3 (in 75-90th percentile) | -27.4% |
| $$$$ | -44.0% | Tier 4 (in 90-99th percentile) | -31.7% |
| $missing | -36.27% | Tier 5 (in 99-100th percentile) | 60.9% |

*Note:* **1** We conduct the same analysis as in Table F3 to produce this table, only changing the restaurant attributes examined. Panel A reports the regression results of equation (2) that examine the advertising effect on the overall CTA conversion rate for restaurants with different attributes. The first column of panel A reports the regression result for restaurants with different price levels, and the second column reports the regression result for restaurants with different volumes of organic page views (traffic) during the three months before the experiment. We rank restaurant review volumes and create tiers based on the percentiles. **2** With the estimates in Panel A and and the discussion in Appendix B.2, Panel B reports the percentage change in CTA conversion rates during the advertising period, comparing page visits from paid clicks and those from organic clicks. Letting $\rho_{org}$ and $\rho_{paid}$ denote the CTA conversion rate of organic and paid page views, respectively, the percentage change in CTA conversion rates (paid vs. organic) measures $(\rho_{paid} - \rho_{org})/\rho_{org}$. **3** The regression is run at the business×month level including the full sample of businesses. It also produces estimates of the heterogeneous effects on the overall CTA conversion rate in the two months after the experiment, and all the coefficients are small and statistically insignificant.

# Appendix G Advertising Effects by Subsample

Each subsample in the data consists of a very different set of restaurants. Part of the objective of the experiment for Yelp is to learn about the heterogeneous advertising effects for each subsample. We discuss how we obtain each subsample in Section 2.2 and show sample sizes in Appendix A Table A1. Overall, the subsample findings are consistent with the heterogeneous advertising effect findings when we cut the full sample by restaurant attributes. We show the results of the advertising effects by subsample in this section as an alternative way to present results, which might be of managerial interest to some readers.

The regression equation to get the treatment effects for each subsample is shown in the following equation:

$$y_{it} = \beta_0 + \sum_{\iota=1}^{2}\sum_{s=2}^{4} \gamma_{s\iota} T_{\iota t} \times S_{si} + \sum_{\iota=1}^{2}\sum_{s=1}^{4} \beta_{s\iota} S_{si}\omega_i T_{\iota t} + \mu_i + \delta_t + \epsilon_{it}. \tag{4}$$

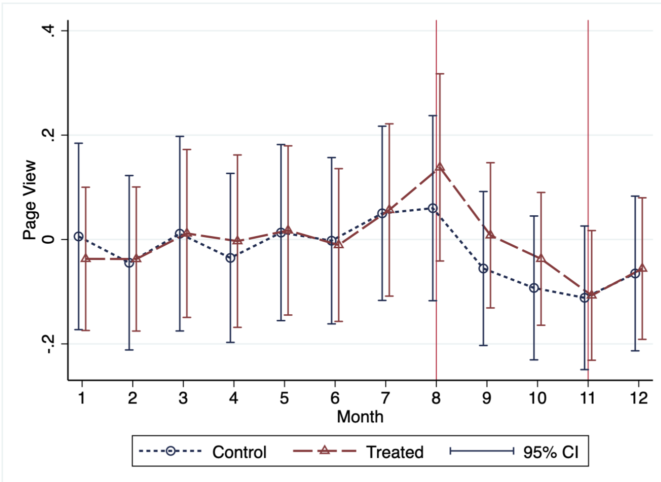
Since the treatment effects are estimated for each subsample separately and are not affected by the fraction of the subsample treated, we no longer need to weight the regressors as we do in equation (1) when trying to get the overall sample ATE. $\beta_{s\iota}$ consistently estimates the treatment effects for subsample $s$ during the experiment when ads are on ($\iota = 1$) and after the experiment when ads are off ($\iota = 1$).

The distribution of restaurant attributes by subsample is shown in Appendix A Table A3. Subsample 1 consists of restaurants that have partnered with Yelp Reservations, and subsample 2 consists of restaurants that have partnered with OpenTable. Restaurants in these two subsamples are mostly full-service restaurants. Compared with other subsamples, restaurants in the first two subsamples are less likely to be chain restaurants, less likely to have low review ratings, and much more likely to be in the top quartile in terms of the number of reviews. Comparing Yelp Reservations restaurants and OpenTable restaurants, Yelp Reservations restaurants are twice as likely to have high review ratings (4.5 or 5 stars) and are much more likely to be new restaurants; hence they are more likely to have accumulated fewer reviews. Given restaurant attributes, we expect advertising effects to be stronger for these restaurants and the conversion rate reduction of paid visits to be lower. Appendix G Table G1 and G2 show the relative size of the effects in different subsamples going in the direction we expected. However, since we only have around 300 Yelp Reservations restaurants, we cannot estimate the effect on this set of restaurants precisely.

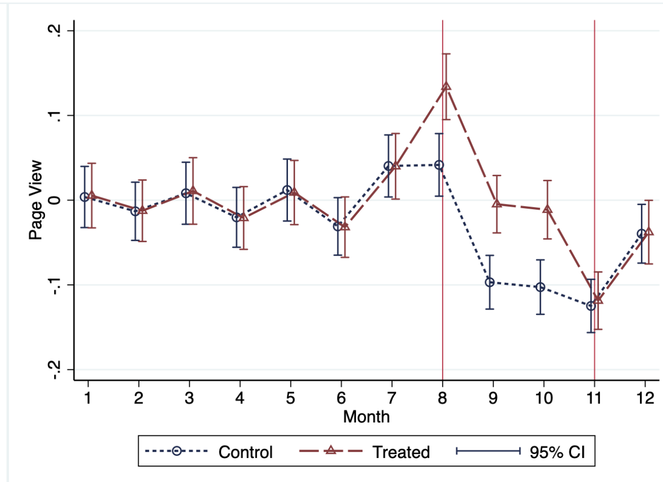
In the main regression, we scale the outcome variables by the same number—the standard deviation in the full sample before the experiment. In this appendix, we also show the size of the effects relative to the standard deviation in each subsample. The results are shown in Appendix G Figures G1–G3. Although the absolute size of the effects is bigger for the first two subsamples, the figures show that subsamples 3 and 4 experience bigger advertising effects relative to their own subsample standard deviations. This is due to bigger standard deviations in the first two subsamples. Similarly, smaller absolute effect sizes do not necessarily mean smaller growth compared with the base level. In the context of different subsamples, the organic traffic for OpenTable restaurants is 2.1 times that of EAT24 restaurants, and the traffic for EAT24 restaurants is 2.7 that of non-partner restaurants. Comparing these base levels with the subsample treatment effects estimated in Appendix G Table G1, the percentage effect is highest for non-partner restaurants, which have the smallest base.

Figure G1: Standardized Subsample Monthly Total Page Views by Treatment Status

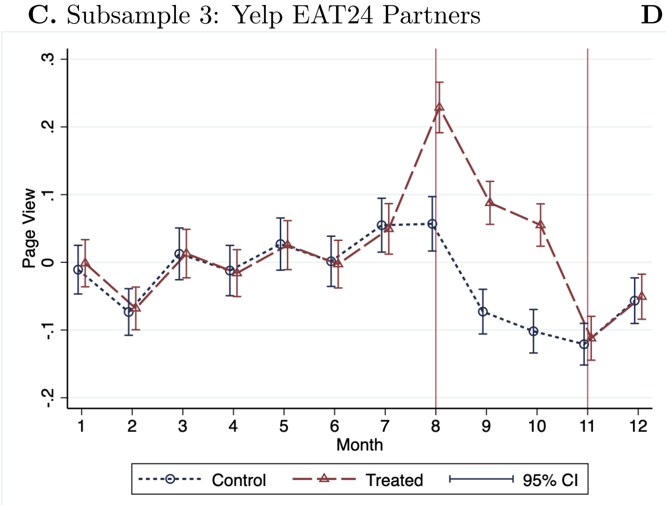
**A.** Subsample 1: Yelp Reservations Partners



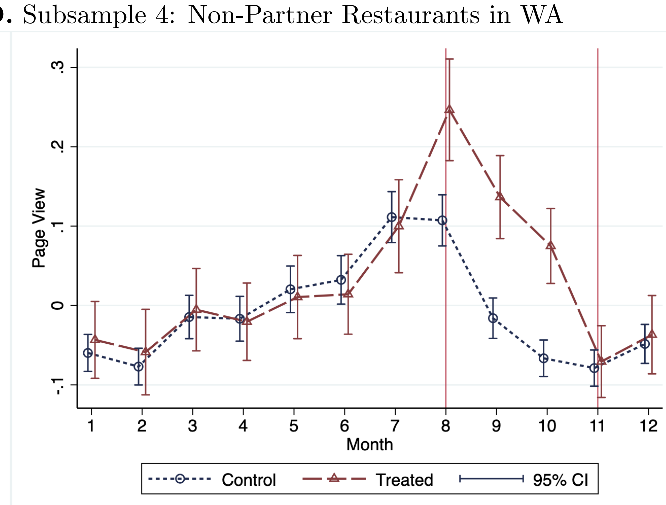

**B.** Subsample 2: OpenTable Partners




**C.** Subsample 3: Yelp EAT24 Partners



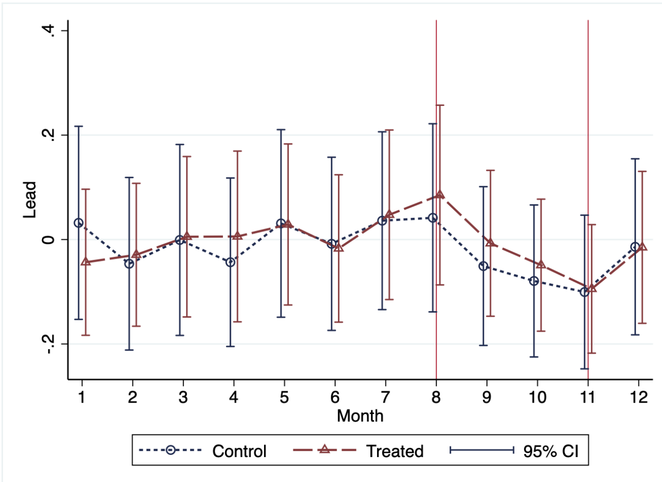

**D.** Subsample 4: Non-Partner Restaurants in WA



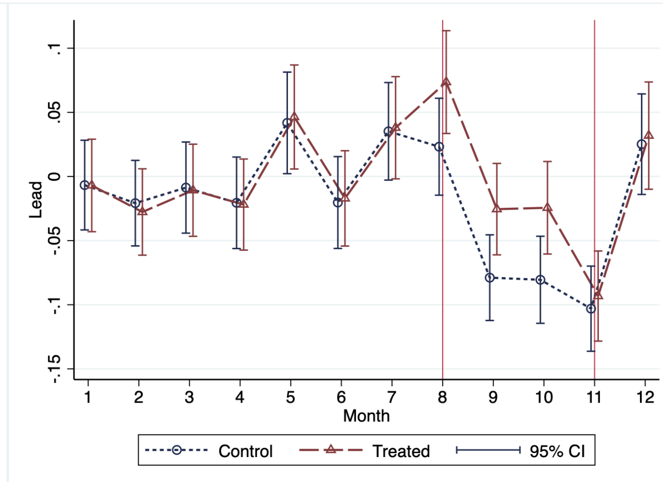

*Notes:* **1** The figures plot the monthly average Yelp page views at the business level. **2** The outcome is standardized by its overall pre-experiment subsample mean and standard deviation, so the effect sizes shown are relative to each subsample's standard deviation. The overall pre-experiment differences in mean outcome between the treated and control sample are taken out. **3** The two solid vertical lines indicate the beginning and end of the experiment. **4** See Appendix A Tables A1 and A3 for information on each subsample.

Figure G2: Standardized Subsample Monthly Total Leads by Treatment Status

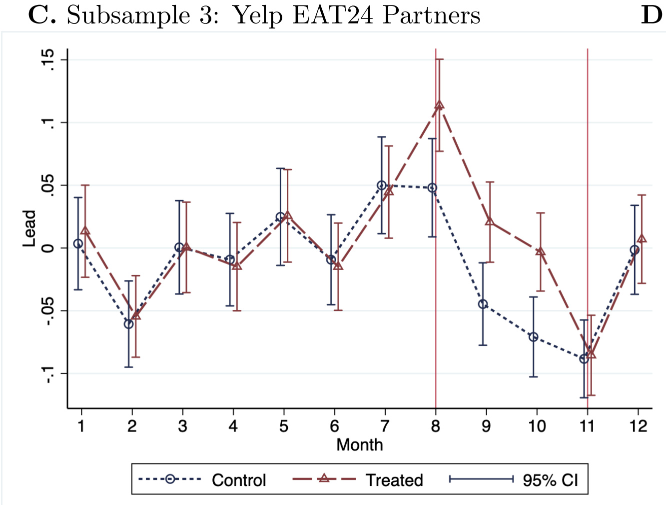
**A.** Subsample 1: Yelp Reservations Partners



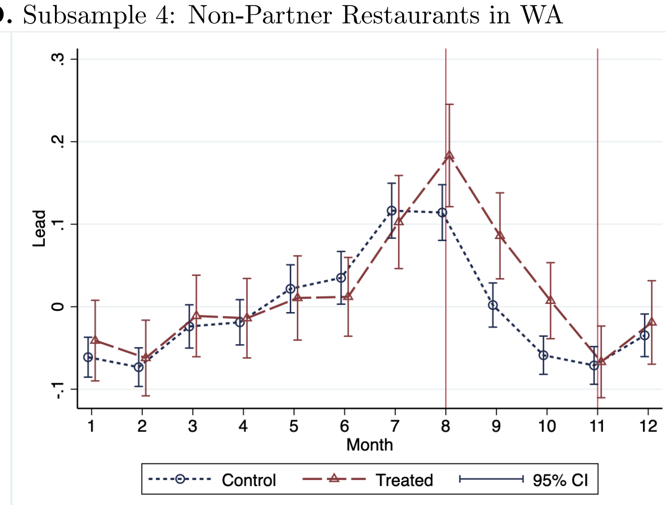

**B.** Subsample 2: OpenTable Partners




**C.** Subsample 3: Yelp EAT24 Partners

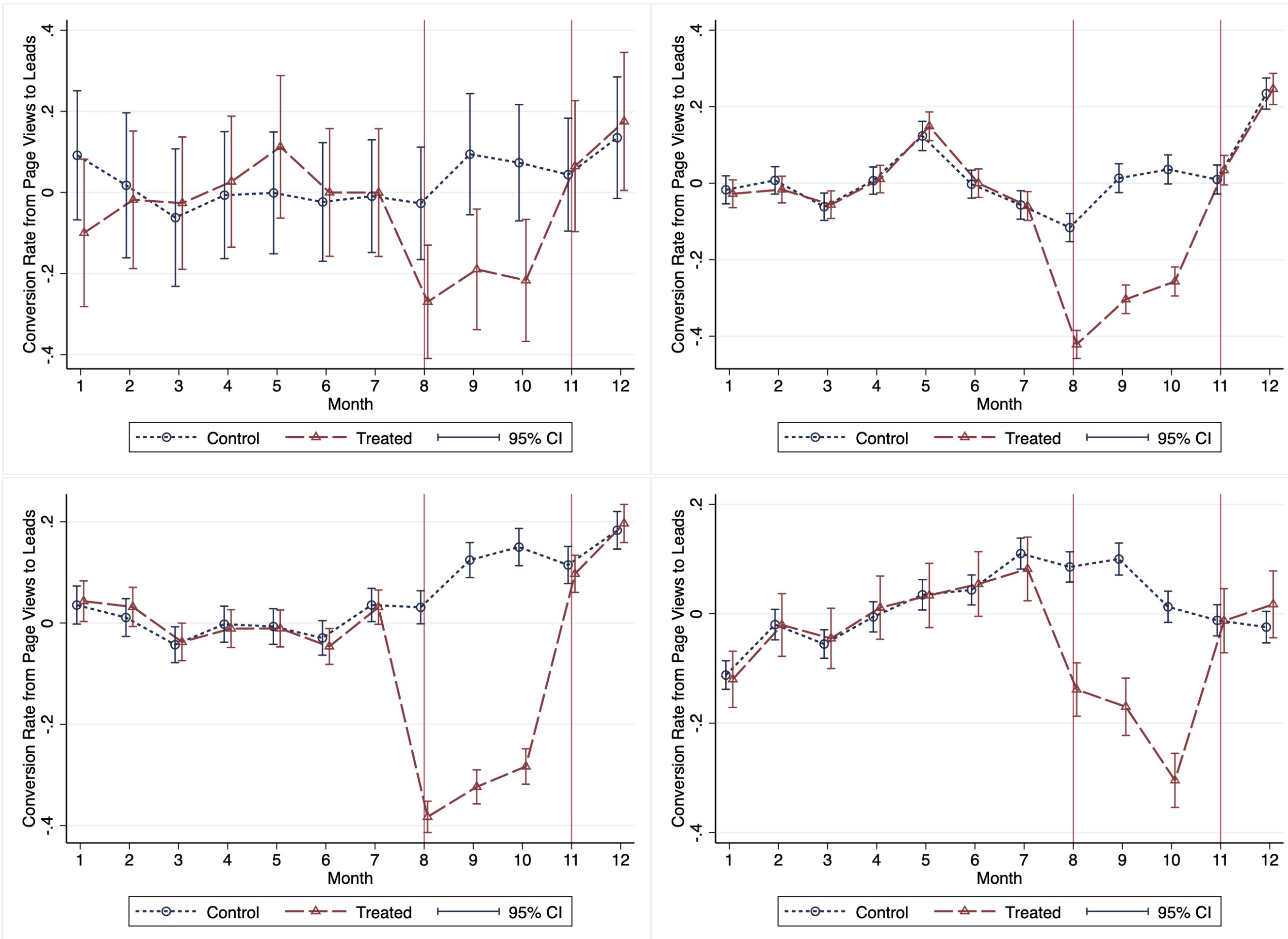



**D.** Subsample 4: Non-Partner Restaurants in WA




*Notes:* **1** The figures plot the monthly leads (requests for direction, calls to restaurants and clicks on the restaurant's external URL) at the business level. **2** The outcome is standardized by its overall pre-experiment subsample mean and standard deviation, so the effect sizes shown are relative to each subsample's standard deviation. The overall pre-experiment differences in mean outcome between the treated and control sample are taken out.**3** The two solid vertical lines indicate the beginning and end of the experiment. **4** See Appendix A Tables A1 and A3 for information on each subsample.

Figure G3: Standardized Subsample Monthly Call-to-Action Conversion Rates by Treatment Status




*Notes:* **1** The figures plot the monthly average call-to-action (CTA) conversion rates at the business level. The CTA conversion rate is calculated by $Leads/PageViews$, measuring the probability of consumers clicking on the CTA buttons (that result in leads) after visiting the restaurant's page. **2** The outcome is standardized by its overall pre-experiment subsample mean and standard deviation, so the effect sizes shown are relative to each subsample's standard deviation. The overall pre-experiment differences in mean outcome between the treated and control sample are taken out. **3** The two solid vertical lines indicate the start and end of the treatment. **4** See Appendix A Tables A1 and A3 for information on each subsample.

Table G1: Average Treatment Effects by Subsample

*[Outcomes Standardized]*

| | (1) Organic Page Views | (2) Total Page Views | (3) Map Views | (4) Calls from Mobile | (5) External URL Clicks | (6) Reviews |
|---|---|---|---|---|---|---|
| Treated×ExprOn×SS1 | 0.018 | 0.145*** | 0.072 | 0.083 | 0.078 | -0.002 |
| | (0.051) | (0.052) | (0.051) | (0.067) | (0.065) | (0.105) |
| Treated×ExprOn×SS2 | 0.001 | 0.125*** | 0.082*** | 0.033*** | 0.07*** | 0.045*** |
| | (0.006) | (0.006) | (0.007) | (0.007) | (0.009) | (0.011) |
| Treated×ExprOn×SS3 | -0.008* | 0.103*** | 0.072*** | 0.042*** | 0.014** | 0.022** |
| | (0.004) | (0.004) | (0.006) | (0.006) | (0.006) | (0.008) |
| Treated×ExprOn×SS4 | -0.006** | 0.051*** | 0.041*** | 0.017*** | 0.011** | 0.017** |
| | (0.002) | (0.003) | (0.004) | (0.003) | (0.004) | (0.006) |
| Treated×PostExpr×SS1 | 0.015 | 0.009 | -0.017 | 0.025 | 0.004 | 0.052 |
| | (0.062) | (0.062) | (0.061) | (0.078) | (0.077) | (0.157) |
| Treated×PostExpr×SS2 | 0.007 | 0.006 | 0.011 | 0.002 | 0.012 | 0.001 |
| | (0.009) | (0.009) | (0.009) | (0.012) | (0.012) | (0.013) |
| Treated×PostExpr×SS3 | 0.007 | 0.005 | 0.012 | 0.002 | -0.002 | -0.005 |
| | (0.006) | (0.006) | (0.008) | (0.01) | (0.007) | (0.01) |
| Treated×PostExpr×SS4 | 0.003 | 0.003 | 0.004 | 0.006 | 0.0004 | 0.008 |
| | (0.003) | (0.003) | (0.004) | (0.004) | (0.004) | (0.006) |
| Other controls & business FE, month FE | x | x | x | x | x | x |
| N | 218,821 | 218,821 | 218,821 | 218,821 | 218,821 | 218,821 |

* $p<0.05$, ** $p<0.01$, *** $p<0.001$. Standard errors in parentheses, clustered at the business level.

*Notes:* The table reports the results from regression equation (4) that estimates the average advertising effects for each subsample during and after the advertising experiment. The coefficients of cross product terms with subsample dummy variables presented in the table represent the treatment effects in each subsample during and after the end of the experiment. They can also be directly compared with each other in showing the relative size of the effects. The observation is at the business×month level, and the outcomes are standardized by the standard deviation of each subsample in the sample period before the experiment (January to July 2015).

Table G2: Average Treatment Effects on Call-to-Action Conversion Rates by Subsample

***Panel A.*** *ATE on Overall CTA Conversion Rates*

| *[Outcomes Standardized]* | |
|---|---|
| | Overall CTA Conversion Rate |
| Treated×ExprOn×SS1 | -0.228*** |
| | (0.046) |
| Treated×ExprOn×SS2 | -0.186*** |
| | (0.008) |
| Treated×ExprOn×SS3 | -0.362*** |
| | (0.013) |
| Treated×ExprOn×SS4 | -0.366*** |
| | (0.023) |
| Treated×PostExpr×SS1 | 0.025 |
| | (0.063) |
| Treated×PostExpr×SS2 | 0.011 |
| | (0.01) |
| Treated×PostExpr×SS3 | -0.001 |
| | (0.016) |
| Treated×PostExpr×SS4 | 0.031 |
| | (0.026) |
| Other controls & business FE, month FE | x |
| N | 218,670 |

* $p$<0.05, ** $p$<0.01, *** $p$<0.001. Standard errors are in parentheses, clustered at the business level.

***Panel B.*** *Effects on CTA Conversion Rates in Percentages*

| During Experiment | % Effects on Overall CTA Conversion Rate | % Loss in CTA Conversion Rate (Paid vs. Organic)[a] |
|---|---|---|
| Subsample1 | -9.1% | -43.3% |
| Subsample2 | -7.6% | -42.9% |
| Subsample3 | -14.0% | -51.5% |
| Subsample4 | -13.4% | -43.4% |

[a] Letting $\rho_{org}$ and $\rho_{paid}$ denote the call-to-action (CTA) conversion rate of organic and paid page views, respectively, the percentage change in the CTA conversion rate (paid vs. organic) measures $(\rho_{paid} - \rho_{org})/\rho_{org}$ during advertising and is calculated following the discussion in Appendix B.2.

*Notes:* **1** The upper panel complements Table C1 and shows the effects on the overall CTA conversion rate by subsample. The analysis is the same as in Table 4, which runs the regression equation (4) on the outcome of the overall CTA conversion rate. **2** The bottom half of the table reports the same effects as reported in the bottom half of Table C1 but shows the effects by subsamples.

# Appendix H Multiple Testing

In section 4.2, we explore the heterogeneity in advertising effects along different dimensions of restaurant attributes (Table 2 - Table 4). In particular, we discretize each attribute and compare the advertising effects between restaurants with different levels of attributes. These comparisons result in a large number of tests that call for multiple testing to account for the error rate of false positive conclusions.

We control the false discovery rate which is defined as the expected proportion of errors among the rejected hypotheses (Benjamini and Hochberg, 1995). We choose the Benjamini-Hochberg procedure in Benjamini and Hochberg (1995) since the hypotheses we test are not independent, and this method produces a conservative p-value cutoff under general dependence structures of the tests (Benjamini and Yekutieli, 2001). Given the theoretical predictions of the directions of the heterogeneous result and to increase the power, we conduct the one-sided tests on the comparison between coefficients.

We test the following hypotheses based on the results in Table 2-4. We test that the advertising effects on page views and leads for restaurants with different levels of attributes are different from zero, and we test restaurants with different levels of attributes within the same dimension (e.g. restaurants with different levels of ratings) have different advertising effects. More specifically, we test whether the advertising effects on page views and leads are higher for restaurants that are not national chains, higher-rated, more reviewed, newer, higher-priced, as well as those that have higher pre-experiment organic traffic. In total, we have 152 tests. We perform the Benjamini-Hochberg procedure, and the adjusted p-value cutoff to reject the null hypothesis is 0.032.

Table H1: P-values of Testing Heterogeneous Advertising Effects by Chain Status

*[Outcomes Standardized]*

| *Weighted Regressors*[a] | (1) Page Views | *p-value* | (2) Leads | *p-value* |
|---|---|---|---|---|
| ExprOn×Treated | | | | |
| ×I(non-chain) $_{(Coeff1)}$ | 0.095*** | | 0.049*** | |
| | (0.003) | $\mathbf{6.8{\times}10^{-217}}$ | (0.003) | $\mathbf{2{\times}10^{-49}}$ |
| ×I(local chain) $_{(Coeff2)}$ | 0.095*** | | 0.053*** | |
| | (0.007) | $\mathbf{6.8{\times}10^{-40}}$ | (0.008) | $\mathbf{1.5{\times}10^{-11}}$ |
| ×I(national chain) $_{(Coeff3)}$ | 0.044*** | | 0.023*** | |
| | (0.004) | $\mathbf{1.6{\times}10^{-24}}$ | (0.004) | $\mathbf{1.2{\times}10^{-8}}$ |
| Other Controls, Fixed Effects | x | | x | |
| N | 218,821 | | 218,821 | |
| **Hypothesis Test ($H_1$)** | *p-value* | | *p-value* | |
| *Coeff1 < Coeff2* | 0.495 | | 0.349 | |
| *Coeff1 < Coeff3* | $\mathbf{1.8{\times}10^{-22}}$ | | $\mathbf{3.3{\times}10^{-7}}$ | |
| *Coeff2 < Coeff3* | $\mathbf{4.9{\times}10^{-10}}$ | | **0.0004** | |

* p<0.05, ** p<0.01, *** p<0.001. Standard errors are in parentheses, clustered at the business level.

*Notes:* **1** The coefficient estimates in the table are copied from Table 2 which shows the advertising effects for restaurants with different chain statuses. **2** Based on the estimates, we test the statistical significance of the estimates, and we test pairwise comparisons of whether non-chain restaurants gain more from advertising. We control the false discovery rate of multiple testing to be under 0.05 using the Benjamini-Hochberg procedure, which yields a p-value cutoff of 0.032 to reject the null hypothesis. We use bold font to indicate tests that are statistically significant after the multiple testing adjustment.

Table H2: P-values of Testing Heterogeneous Advertising Effects by Review Rating
*[Outcomes Standardized]*

| *Weighted Regressors* | (1) Page Views | *p-value* | (2) Leads | *p-value* |
|---|---|---|---|---|
| ExprOn×Treated | | | | |
| ×I(3 Stars) $_{(Coeff1)}$ | 0.073*** | | 0.021*** | |
| | (0.004) | $\mathbf{4.5{\times}10^{-83}}$ | (0.003) | $\mathbf{5.1{\times}10^{-11}}$ |
| ×I(3.5 Stars) $_{(Coeff2)}$ | 0.092*** | | 0.041*** | |
| | (0.004) | $\mathbf{4.4{\times}10^{-125}}$ | (0.004) | $\mathbf{3.4{\times}10^{-30}}$ |
| ×I(4 Star) $_{(Coeff3)}$ | 0.101*** | | 0.062*** | |
| | (0.005) | $\mathbf{3.2{\times}10^{-88}}$ | (0.006) | $\mathbf{1.2{\times}10^{-26}}$ |
| ×I(>4 Star) $_{(Coeff4)}$ | 0.097*** | | 0.062*** | |
| | (0.010) | $\mathbf{2.7{\times}10^{-21}}$ | (0.013) | $\mathbf{1.2{\times}10^{-6}}$ |
| Other Controls, Fixed Effects | x | | x | |
| N | 216,586 | | 216,586 | |
| **Hypothesis Test ($H_1$)** | *p-value* | | *p-value* | |
| *Coeff1 < Coeff2* | $\mathbf{4.40{\times}10^{-5}}$ | | $\mathbf{1.28{\times}10^{-5}}$ | |
| *Coeff1 < Coeff3* | $\mathbf{8.53{\times}10^{-7}}$ | | $\mathbf{2.37{\times}10^{-10}}$ | |
| *Coeff1 < Coeff4* | **0.008** | | **0.001** | |
| *Coeff2 < Coeff3* | 0.077 | | **0.001** | |
| *Coeff2 < Coeff4* | 0.308 | | 0.061 | |
| *Coeff3 < Coeff4* | 0.378 | | 0.480 | |

* p<0.05, ** p<0.01, *** p<0.001. Standard errors are in parentheses, clustered at the business level.

*Notes:* **1** The coefficient estimates in the table are copied from Table 2 which shows the advertising effects for restaurants with different average review ratings. **2** Based on the estimates, we test the statistical significance of the estimates, and we test pairwise comparisons of whether restaurants with higher review ratings gain more from advertising. We control the false discovery rate of multiple testing to be under 0.05 using the Benjamini-Hochberg procedure, which yields a p-value cutoff of 0.032 to reject the null hypothesis. We use bold font to indicate tests that are statistically significant after the multiple testing adjustment.

Table H3: P-values of Testing Heterogeneous Advertising Effects by Baseline Organic Traffic

| *[Outcomes Standardized]* | | | | |
|---|---|---|---|---|
| *Weighted Regressors*[a] | (1) Page Views | *p-value* | (2) Leads | *p-value* |
| ExprOn×Treated | | | | |
| ×I(OrgnicTraffic≤25th pctl) $_{(Coeff1)}$ | 0.054*** | | 0.022*** | |
| | (0.001) | **$<10^{-323}$** | (0.001) | **$1.4\times10^{-80}$** |
| ×I(OrgnicTraffic∈(25th,50th] pctl) $_{(Coeff2)}$ | 0.087*** | | 0.034*** | |
| | (0.002) | **$<10^{-323}$** | (0.002) | **$4.4\times10^{-48}$** |
| ×I(OrgnicTraffic∈(50th,75th] pctl) $_{(Coeff3)}$ | 0.106*** | | 0.052*** | |
| | (0.004) | **$3.2\times10^{-192}$** | (0.004) | **$3.2\times10^{-45}$** |
| ×I(OrgnicTraffic∈(75th,90th] pctl) $_{(Coeff4)}$ | 0.107*** | | 0.065*** | |
| | (0.007) | **$2.7\times10^{-48}$** | (0.008) | **$2.7\times10^{-15}$** |
| ×I(OrgnicTraffic<90th) pctl) $_{(Coeff5)}$ | 0.133*** | | 0.102*** | |
| | (0.021) | **$2.7\times10^{-10}$** | (0.024) | **$2.8\times10^{-5}$** |
| Other Controls, Fixed Effects | x | | x | |
| N | 218,821 | | 218,821 | |
| **Hypothesis Test ($H_1$)** | *p-value* | | *p-value* | |
| *Coeff1 < Coeff2* | **$2.0\times10^{-37}$** | | **$1.2\times10^{-6}$** | |
| *Coeff1 < Coeff3* | **$1.0\times10^{-42}$** | | **$6.1\times10^{-15}$** | |
| *Coeff1 < Coeff4* | **$4.9\times10^{-13}$** | | **$1.3\times10^{-7}$** | |
| *Coeff1 < Coeff5* | **0.0001** | | **0.0005** | |
| *Coeff2 < Coeff3* | **$2.0\times10^{-6}$** | | **$3.8\times10^{-5}$** | |
| *Coeff2 < Coeff4* | **0.004** | | **0.0002** | |
| *Coeff2 < Coeff5* | **0.016** | | **0.003** | |
| *Coeff3 < Coeff4* | 0.450 | | 0.067 | |
| *Coeff3 < Coeff5* | 0.105 | | **0.020** | |
| *Coeff4 < Coeff5* | 0.124 | | 0.075 | |

* p<0.05, ** p<0.01, *** p<0.001. Standard errors are in parentheses, clustered at the business level.

*Notes:* **1** The coefficient estimates in the table are copied from Table 3 which shows the advertising effects for restaurants with different pre-experiment organic traffic levels. **2** Based on the estimates, we test the statistical significance of the estimates, and we test pairwise comparisons of whether restaurants with higher pre-experiment organic traffic gain more from advertising. We control the false discovery rate of multiple testing to be under 0.05 using the Benjamini-Hochberg procedure, which yields a p-value cutoff of 0.032 to reject the null hypothesis. We use bold font to indicate tests that are statistically significant after the multiple testing adjustment.

Table H4: P-values of Testing Heterogeneous Advertising Effects by Number of Reviews
*[Outcomes Standardized]*

| *Weighted Regressors* | (1) Page Views | *p-value* | (2) Leads | *p-value* |
|---|---|---|---|---|
| $\times$(#Rev$\leq$25th pctl) $_{(Coeff1)}$ | 0.061*** | | 0.027*** | |
| | (0.002) | $\mathbf{6.6{\times}10^{-141}}$ | (0.002) | $\mathbf{5.6{\times}10^{-37}}$ |
| $\times$(#Rev$\in$(25th,50th] pctl) $_{(Coeff2)}$ | 0.089*** | | 0.039*** | |
| | (0.004) | $\mathbf{2.7{\times}10^{-123}}$ | (0.004) | $\mathbf{3.7{\times}10^{-23}}$ |
| $\times$(#Rev$\in$(50th,75th] pctl) $_{(Coeff3)}$ | 0.102*** | | 0.049*** | |
| | (0.005) | $\mathbf{1.7{\times}10^{-91}}$ | (0.006) | $\mathbf{8.8{\times}10^{-18}}$ |
| $\times$(#Rev$\in$(75th,90th] pctl) $_{(Coeff4)}$ | 0.114*** | | 0.066*** | |
| | (0.008) | $\mathbf{1.4{\times}10^{-46}}$ | (0.009) | $\mathbf{2.3{\times}10^{-14}}$ |
| $\times$(#Rev>90th pctl) $_{(Coeff5)}$ | 0.116*** | | 0.086*** | |
| | (0.017) | $\mathbf{2.0{\times}10^{-12}}$ | (0.019) | $\mathbf{6.3{\times}10^{-6}}$ |
| Other Controls, Fixed Effects | x | | x | |
| N | 218,821 | | 218,821 | |
| **Hypothesis Test ($H_1$)** | *p-value* | | *p-value* | |
| *Coeff1 < Coeff2* | $\mathbf{1.28{\times}10^{-11}}$ | | **0.0027** | |
| *Coeff1 < Coeff3* | $\mathbf{4.15{\times}10^{-14}}$ | | $\mathbf{8.95{\times}10^{-5}}$ | |
| *Coeff1 < Coeff4* | $\mathbf{8.82{\times}10^{-11}}$ | | $\mathbf{4.21{\times}10^{-6}}$ | |
| *Coeff1 < Coeff5* | **0.0005** | | **0.001** | |
| *Coeff2 < Coeff3* | **0.021** | | 0.071 | |
| *Coeff2 < Coeff4* | **0.003** | | **0.002** | |
| *Coeff2 < Coeff5* | 0.054 | | **0.008** | |
| *Coeff3 < Coeff4* | 0.104 | | 0.049 | |
| *Coeff3 < Coeff5* | 0.200 | | **0.032** | |
| *Coeff4 < Coeff5* | 0.440 | | 0.176 | |

* p<0.05, ** p<0.01, *** p<0.001. Standard errors are in parentheses, clustered at the business level.
*Notes:* **1** The coefficient estimates in the table are copied from Table 3 which shows the advertising effects for restaurants with different numbers of reviews. **2** Based on the estimates, we test the statistical significance of the estimates, and we test pairwise comparisons of whether restaurants with more reviews gain more from advertising. We control the false discovery rate of multiple testing to be under 0.05 using the Benjamini-Hochberg procedure, which yields a p-value cutoff of 0.032 to reject the null hypothesis. We use bold font to indicate tests that are statistically significant after the multiple testing adjustment.

Table H5: P-values of Testing Heterogeneous Advertising Effects by Restaurant Age
*[Outcomes Standardized]*

| *Weighted* | (1) | | (2) | |
|---|---|---|---|---|
| *Regressors*[a] | Page Views | *p-value* | Leads | *p-value* |
| ExprOn×Treated | | | | |
| ×I(age<1yr) $_{(Coeff1)}$ | 0.119*** | | 0.069** | |
| | (0.023) | $\mathbf{2.5{\times}10^{-7}}$ | (0.024) | **0.004** |
| ×I(age∈[1,2]yr) $_{(Coeff2)}$ | 0.103*** | | 0.066*** | |
| | (0.013) | $\mathbf{7.8{\times}10^{-16}}$ | (0.015) | $\mathbf{7.8{\times}10^{-6}}$ |
| ×I(age∈[3,4]yr) $_{(Coeff3)}$ | 0.090*** | | 0.032*** | |
| | (0.007) | $\mathbf{2.2{\times}10^{-36}}$ | (0.007) | $\mathbf{1.6{\times}10^{-5}}$ |
| ×I(age∈[5,10]yr) $_{(Coeff4)}$ | 0.097*** | | 0.054*** | |
| | (0.004) | $\mathbf{1.0{\times}10^{-149}}$ | (0.004) | $\mathbf{1.2{\times}10^{-33}}$ |
| ×I(age>10yr) $_{(Coeff5)}$ | 0.081*** | | 0.040*** | |
| | (0.003) | $\mathbf{1.3{\times}10^{-143}}$ | (0.003) | $\mathbf{8.4{\times}10^{-35}}$ |
| Other Controls,Fixed Effects | x | | x | |
| N | 218,266 | | 218,266 | |
| **Hypothesis Test ($H_1$)** | *p-value* | | *p-value* | |
| *Coeff1 < Coeff2* | 0.265 | | 0.460 | |
| *Coeff1 < Coeff3* | 0.112 | | 0.071 | |
| *Coeff1 < Coeff4* | 0.167 | | 0.275 | |
| *Coeff1 < Coeff5* | 0.049 | | 0.120 | |
| *Coeff2 < Coeff3* | 0.188 | | **0.019** | |
| *Coeff2 < Coeff4* | 0.325 | | 0.221 | |
| *Coeff2 < Coeff5* | 0.045 | | 0.045 | |
| *Coeff3 < Coeff4* | 0.195 | | **0.004** | |
| *Coeff3 < Coeff5* | 0.117 | | 0.143 | |
| *Coeff4 < Coeff5* | **0.0004** | | **0.006** | |

* p<0.05, ** p<0.01, *** p<0.001. Standard errors are in parentheses, clustered at the business level.

*Notes:* **1** The coefficient estimates in the table are copied from Table 4 which shows the advertising effects for restaurants of different ages. **2** Based on the estimates, we test the statistical significance of the estimates, and we test pairwise comparisons of whether new restaurants gain more from advertising. We control the false discovery rate of multiple testing to be under 0.05 using the Benjamini-Hochberg procedure, which yields a p-value cutoff of 0.032 to reject the null hypothesis. We use bold font to indicate tests that are statistically significant after the multiple testing adjustment.

Table H6: P-values of Testing Heterogeneous Advertising Effects by Price Category
*[Outcomes Standardized]*

| *Weighted* *Regressors*[a] | (1) Page Views | *p-value* | (2) Leads | *p-value* |
|---|---|---|---|---|
| ExprOn×Treated | | | | |
| ×I(missing) $_{(Coeff1)}$ | 0.047*** | | 0.019*** | |
| | (0.011) | $\mathbf{5.0\times10^{-159}}$ | (0.006) | $\mathbf{1.0\times10^{-33}}$ |
| ×I($) $_{(Coeff2)}$ | 0.078*** | | 0.038*** | |
| | (0.003) | $\mathbf{1.3\times10^{-136}}$ | (0.003) | $\mathbf{2.5\times10^{-31}}$ |
| ×I($$) $_{(Coeff3)}$ | 0.095*** | | 0.051*** | |
| | (0.004) | $\mathbf{2.2\times10^{-28}}$ | (0.004) | $\mathbf{5.8\times10^{-8}}$ |
| ×I($$$) $_{(Coeff4)}$ | 0.120*** | | 0.064*** | |
| | (0.011) | **0.0009** | (0.012) | 0.39 |
| ×I($$$$) $_{(Coeff5)}$ | 0.097*** | | 0.022 | |
| | (0.029) | $\mathbf{9.2\times10^{-6}}$ | (0.025) | **0.0006** |
| Other Controls, Fixed Effects | x | | x | |
| N | 218,821 | | 218,821 | |
| **Hypothesis Test ($H_1$)** | *p-value* | | *p-value* | |
| *Coeff1 < Coeff2* | **0.0003** | | **0.008** | |
| *Coeff1 < Coeff3* | **0.0002** | | **0.020** | |
| *Coeff1 < Coeff4* | 0.264 | | 0.262 | |
| *Coeff1 < Coeff5* | **0.006** | | **0.0003** | |
| *Coeff2 < Coeff3* | **0.019** | | 0.171 | |
| *Coeff2 < Coeff4* | 0.472 | | 0.127 | |
| *Coeff2 < Coeff5* | $\mathbf{5\times10^{-5}}$ | | $\mathbf{2.7\times10^{-7}}$ | |
| *Coeff3 < Coeff4* | 0.233 | | 0.070 | |
| *Coeff3 < Coeff5* | $\mathbf{4.7\times10^{-6}}$ | | **0.0002** | |
| *Coeff4 < Coeff5* | 0.069 | | 0.421 | |

* $p<0.05$, ** $p<0.01$, *** $p<0.001$. Standard errors are in parentheses, clustered at the business level.
*Notes:* **1** The coefficient estimates in the table are copied from Table 4 which shows the advertising effects for restaurants with different price levels. **2** Based on the estimates, we test the statistical significance of the estimates, and we test pairwise comparisons of whether pricier restaurants gain more from advertising. We control the false discovery rate of multiple testing to be under 0.05 using the Benjamini-Hochberg procedure, which yields a p-value cutoff of 0.032 to reject the null hypothesis. We use bold font to indicate tests that are statistically significant after the multiple testing adjustment.

## References


Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1), 289-300.

Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of statistics*, 1165-1188.